cemmap

centre for microdata methods and practice

Nonparametric tests of conditional treatment effects

Sokbae Lee
Yoon-Jae Whang

The Institute for Fiscal Studies
Department of Economics, UCL

cemmap working paper CWP36/09

An ESRC Research Centre

Institute for Fiscal Studies, 7 Ridgmount Street, London WC1E 7AE **Tel:** 44 (0)20 7291 4800 **Email:** info@cemmap.ac.uk **www.cemmap.ac.uk**

NONPARAMETRIC TESTS OF CONDITIONAL TREATMENT EFFECTS

SOKBAE LEE AND YOON-JAE WHANG

Abstract. We develop a general class of nonparametric tests for treatment effects conditional on covariates. We consider a wide spectrum of null and alternative hypotheses regarding conditional treatment effects, including (i) the null hypothesis of the conditional stochastic dominance between treatment and control groups; (ii) the null hypothesis that the conditional average treatment effect is positive for each value of covariates; and (iii) the null hypothesis of no distributional (or average) treatment effect conditional on covariates against a one-sided (or two-sided) alternative hypothesis. The test statistics are based on L_1-type functionals of uniformly consistent nonparametric kernel estimators of conditional expectations that characterize the null hypotheses. Using the Poissionization technique of Giné et al. (2003), we show that suitably studentized versions of our test statistics are asymptotically standard normal under the null hypotheses and also show that the proposed nonparametric tests are consistent against general fixed alternatives. Furthermore, it turns out that our tests have non-negligible powers against some local alternatives that are $n^{-1/2}$ different from the null hypotheses, where n is the sample size. We provide a more powerful test for the case when the null hypothesis may be binding only on a strict subset of the support and also consider an extension to testing for quantile treatment effects. We illustrate the usefulness of our tests by applying them to data from a randomized, job training program (LaLonde, 1986) and by carrying out Monte Carlo experiments based on this dataset.

Key words. Average treatment effect, conditional stochastic dominance, Poissionization, programme evaluation.

JEL Subject Classification. C12, C14, C21.

AMS Subject Classification. 62G10, 62G20.

Date: 2 November 2009.
We would like to thank Donald Andrews, Herman Bierens, Xiahong Chen, Yanqin Fan, Joel Horowitz, Hidehiko Ichimura, Yuichi Kitamura, Tong Li, Oliver Linton, David Mason, Whitney Newey, James Powell, Edward Vytlacil, Quang Vuong, and participants at several seminars and conferences for useful suggestions and helpful comments. Lee thanks the Economic and Social Research Council for the ESRC Centre for Microdata Methods and Practice (RES-589-28-0001) and the small research grant (RES-000-22-2761) and the European Research Council for the research grant (ERC-2009-StG-240910-ROMETA). Lee also thanks the Cowles Foundation for Research in Economics at Yale University, where he is currently a visiting fellow. Whang thanks the Cowles Foundation for Research in Economics at Yale University and the Centre for Microdata Methods and Practice at University College London for their hospitality during his visits.

1. Introduction

Recent years have witnessed a surge of applied research using data from random assignment of treatment to a social program as an attempt to provide a credible answer to important economic questions.[1] Randomized programme evaluation is a part of a much larger literature on econometric evaluation of social programs. For recent reviews of this huge literature, see, e.g., Abbring and Heckman (2007), Blundell and Costa Dias (2008), Heckman and Vytlacil (2007a,b), Imbens (2004), and Imbens and Wooldridge (2009), among others. Most of the literature focused on point or set identification of treatment effect parameters, on estimation of identified parameters, and also on relevance of randomized experiments. However, there has been much less attention devoted to testing hypotheses regarding treatment effects. This might be due to the fact that typically the main focus of empirical work has been on estimation of average treatment effects for the entire population or for the treated.[2] For these parameters, standard inference can be applied to test the null hypothesis of no average treatment effect. However, one ubiquitous feature of treatment effects in the program evaluation literature is that treatment effects tend to vary across different groups and individuals. Also, average treatment effects might not provide a full picture of treatment effects since it is possible to have significant distributional treatment effects at the top or bottom of the population distribution with zero average treatment effect. For recent empirical evidence on importance of distributional treatment effects, see, e.g. Bitler et al. (2006, 2007). Therefore, there are other interesting hypotheses to consider, as emphasized in Imbens and Wooldridge (2009, Section 3.3).

In this paper we develop nonparametric tests for both average and distributional treatment effects conditional on covariates. We consider a wide spectrum of null and alternative hypotheses regarding conditional treatment effects, including (i) the null hypothesis of the conditional stochastic dominance between treatment and control groups; (ii) the null hypothesis that the conditional average treatment effect is positive for each value of covariates; and (iii) the null hypothesis of no distributional (or average) treatment effect conditional on covariates against a one-sided (or two-sided) alternative hypothesis.

Although there exists a very large literature on treatments effects and program evaluation, there seem to be only a few related papers in the literature. Abadie (2002) considered

[1]A few recent examples include: the effects of deworming on health and education with school-based mass treatment in rural Kenya (Miguel and Kremer, 2004); the impact of women's leadership on policy decisions using a unique experiment implemented in India (Chattopadhyay and Duflo, 2004); the experimental impacts of the earnings subsidy in a Canadian welfare program (Card and Hyslop, 2005); investigation of neighborhood effects based on social experiments using randomized housing vouchers in U.S. cities (Kling et al., 2007), among many others. See also Glenn and List (2004) for a survey of field experiments and Duflo et al. (2007) for a review on experimental methods in development economics.

[2] For estimation of average treatment effects, see Abadie and Imbens (2006); Chen et al. (2008); Hahn (1998); Heckman and Todd (1998); Hirano et al. (2003), among others.

the null hypotheses of the equality and first-order stochastic dominance between treatment and control groups and developed bootstrap tests. In his setup, there are no covariates and hence, there is no treatment effect heterogeneity by covariates. Linton and Gozalo (1997) considered testing for the conditional independence, mentioning that the null hypothesis of no average treatment effect as an example. Angrist and Kuersteiner (2004, 2008) developed a semiparametric test for conditional independence in time series models with a binary or multinomial policy variable. Crump et al. (2008) developed nonparametric tests for the treatment effect heterogeneity. In particular, they proposed series-estimation-based tests for the null hypothesis that the treatment has a zero average effect for each value of covariates and also for the null hypothesis that the average effect conditional on the covariates is constant. Lee (2009) developed a nonparametric test of the null hypothesis of no distributional treatment effect for randomly censored outcomes.

Except for Abadie (2002), all the null hypotheses considered in the literature are based on equality between functionals of the distributions of treatment and control groups. They are relatively easier to deal with using the existing statistical tools.[3] However, some of our null hypotheses of interest are based on inequality between functionals of the distributions of treatment and control groups. One such example is the conditional stochastic dominance.[4] Testing conditional stochastic dominance is important beyond the treatment effect setup. For example, in auction theory, Guerre et al. (2009) show that there are testable stochastic dominance relations among observed bid distributions if participation is exogenous. As mentioned in Guerre et al. (2009), if bidders' participation is independent of bidders' private values only after conditioning on a vector of covariates, then it is essential to consider conditional stochastic dominance to test the implications of auction theory.

Our proposed statistics are based on L_1-type functionals of uniformly consistent nonparametric kernel estimators of conditional expectations that characterize the null hypotheses. For example, testing the null of zero conditional average treatment effect against an alterative of positive treatment effect for some values of covariates involves a test statistic such as $\int_{-\infty}^{\infty} \max\{\hat{\tau}(x), 0\} w(x) dx$, where $\hat{\tau}(x)$ is a nonparametric estimator of the conditional average treatment effect and $w(x)$ is a weight function. Testing the same null hypothesis against an alterative of nonzero treatment effect for some values of covariates can be carried out

[3]These exists a substantial literature for testing equality between conditional mean functions or between conditional distribution functions. For example, see Delgado and González Manteiga (2001), Lavergne (2001), Su and White (2004, 2007, 2008), and Song (2007) among others.

[4]There is a large literature on stochastic dominance without covariates: see McFadden (1989), Klecan et al. (1991), Kaur et al. (1994), Anderson (1996), Davidson and Duclos (1997, 2000), Barrett and Donald (2003), Linton et al. (2005), Horváth et al. (2006), Linton et al. (2010), among others. To our best knowledge, there is no test available for stochastic dominance conditional on continuous covariates, which has been an important, open question in the literature.

using a test statistic such as $\int_{-\infty}^{\infty} |\hat{\tau}(x)| w(x) dx$. The exact form of the test statistics varies depending on the type of the null and alternative hypotheses.

To deal with both equality- and inequality-involving null hypotheses, we develop unified asymptotic theory based on the Poissionization technique of Giné et al. (2003). Our theory covers test statistics of L_1-type forms: e.g. $\int_{-\infty}^{\infty} \max\{\hat{\tau}(x), 0\} w(x) dx$ and $\int_{-\infty}^{\infty} |\hat{\tau}(x)| w(x) dx$. In particular, we show that suitably studentized versions of our test statistics are asymptotically standard normal under the null hypotheses and also show that the proposed nonparametric tests are consistent against general fixed alternatives. Furthermore, it turns out that our tests have non-negligible powers against some, though not all, local alternatives that are $n^{-1/2}$ different from the null hypotheses, where n is the sample size. This suggests that for the null hypothesis of zero conditional average effect, our test could be more powerful in some directions than that of Crump et al. (2008) for sufficiently large n, since their test cannot detect $n^{-1/2}$ alternatives. The asymptotic normality with the $n^{-1/2}$ consistency for L_1-type functionals are powerful new results and can be of independent interest.

The remainder of the paper is organized as follows. Section 2 gives our testing framework in the context of program evaluation and Section 3 provides a description of our test statistics. Section 4 establishes asymptotic theory for our test statistics both when the null hypothesis is true and when it is false. We provide some informal description of our proof technique and discuss the choice of the weight function. In the case of the null hypothesis that is expressed in terms of inequality restrictions, in Section 5 we show that we can improve the power performance of our test by estimating the "contact set" on which the inequality restriction is binding. In this section, we also provide test statistics for quantile treatment effects. Section 6 illustrates the usefulness of our testing method by applying it to data from a randomized, job training program (LaLonde, 1986) and by carrying out Monte Carlo experiments based on this dataset. Section 7 gives some concluding remarks. Appendix contains all the proofs of theorems given in the main text.

2. Testing for Conditional Treatment Effects in Program Evaluation

In this section, we describe our hypothesis testing problem in the context of program evaluation. Let Y_1 and Y_0 be potential individual outcomes in two states, with treatment and without treatment. For each individual, the observed outcome Y is $Y = D \cdot Y_1 + (1 - D) \cdot Y_0$, where D denotes an indicator variable for the treatment, with $D = 0$ if an individual is not treated and $D = 1$ if an individual is treated. We assume that independent and identically distributed observations $\{(Y_i, D_i, X_i) : i = 1, \ldots, n\}$ of (Y, D, X) are available, where X denotes a vector of covariates. Let $\mathcal{Y} \times \mathcal{X}$ denote the support of (Y, X).

To describe our null hypotheses of interest, let $G(Y_j, y)$ be a measurable, known function of Y_j with an index y for $j = 0, 1$. The first class of tests is concerned with the null hypothesis

$$\mathcal{H}_0 : E[G(Y_1, y) - G(Y_0, y)|X = x] \leq 0 \;\text{ for each } (y, x) \in \mathcal{W} \tag{2.1}$$

against the alternative hypothesis

$$\mathcal{H}_1 : E[G(Y_1, y) - G(Y_0, y)|X = x] > 0 \;\text{ for some } (y, x) \in \mathcal{W}, \tag{2.2}$$

where $\mathcal{W} := \mathcal{W}_y \times \mathcal{W}_x$ denotes a subset of $\mathcal{Y} \times \mathcal{X}$ on which one wishes to evaluate the treatment effect. The hypothesis (2.1) is a strong hypothesis since it needs to hold for *all* values of (y, x) in $\mathcal{W}$, but can be reduced in strength by limiting $\mathcal{W}$ for which (2.1) holds.

Testing (2.1) is of interest in a number of settings in program evaluation. For example, if $G(Y_j, y) \equiv -Y_j$ for $j = 0, 1$, testing (2.1) amounts to testing the null hypothesis that the conditional average treatment effect is positive for each $x \in \mathcal{W}_x$. If $G(Y_j, y) = 1(Y_j \leq y)$ for $j = 0, 1$, then testing (2.1) amounts to testing the conditional stochastic dominance between treatment and control groups.

The second class of tests is concerned with the null hypothesis

$$\mathcal{H}_0^D : E[G(Y_1, y) - G(Y_0, y)|X = x] = 0 \;\text{ for each } (y, x) \in \mathcal{W} \tag{2.3}$$

against the alternative hypothesis

$$\mathcal{H}_1^D : E[G(Y_1, y) - G(Y_0, y)|X = x] \neq 0 \;\text{ for some } (y, x) \in \mathcal{W}. \tag{2.4}$$

When $G(Y_j, y) \equiv Y_j$ for $j = 0, 1$, the null hypothesis (2.3) is previously considered in Crump et al. (2008). When $G(Y_j, y) = 1(Y_j \leq y)$ for $j = 0, 1$, we test the null hypothesis of equality between conditional distributions between treatment and control groups. This hypothesis is mentioned as an interesting hypothesis to consider by Imbens and Wooldridge (2009, Section 5.12).

In addition, one may consider testing (2.3) against a one-sided alternative such as (2.2). For example, if $G(Y_j, y) \equiv Y_j$ for $j = 0, 1$, testing (2.3) against (2.2) amounts to testing the null hypothesis that the conditional average treatment effect is zero for each $x \in \mathcal{W}_x$ against the alternative hypothesis that the conditional average treatment effect is positive for some $x \in \mathcal{W}_x$. To carry out this, we can use our first class of tests by restricting the null hypothesis in (2.1) to be the least favorable case.[5]

In general, treatment effects are evaluated in three setups: one under randomized experiments, another under the unconfoundedness assumption, and the third under selection on unobservables. For the first two setups and a particular case of the third setup, we develop hypothesis testing for treatment effects conditional on covariates. Suppose that we have data

[5] Thus, in this paper, we will not develop tests of (2.3) against (2.2) separately.

available on random assignment of treatment to a social program. In this case, note that

$$E[G(Y_1, y) - G(Y_0, y)|X = x] = E[G(Y, y)|X = x, D = 1] - E[G(Y, y)|X = x, D = 0]. \tag{2.5}$$

Hence, our test of the null hypothesis (2.1) can be carried out by testing the null hypothesis

$$H_0 : E[G(Y, y)|X = x, Z = 1] \leq E[G(Y, y)|X = x, Z = 0] \ \text{ for each } (y, x) \in \mathcal{W}, \tag{2.6}$$

provided that the following standard overlap assumption is satisfied:

$$0 < \Pr(D = 1|X = x) < 1 \ \text{ for all } x \in \mathcal{W}_x. \tag{2.7}$$

Similarly, a test of (2.3) can be accomplished by testing the null hypothesis

$$H_0^D : E[G(Y, y)|X = x, Z = 1] = E[G(Y, y)|X = x, Z = 0] \ \text{ for each } (y, x) \in \mathcal{W}, \tag{2.8}$$

assuming that (2.7) holds. We are not aware of any existing test that can carry out testing the null hypothesis (2.6). The test of (2.8) can be viewed as testing for significance of Z in $E[Y|X, Z]$ when $G(Y, y) \equiv Y$ and also can be regarded as testing for conditional independence between Y and Z given X when $G(Y, y) = 1(Y \leq y)$. Existing tests of (2.8) typically use the L_2 norm. Using L_1-type functionals, we provide new test statistics for testing (2.8).

For the second setup, the unconfoundedness assumption, that is Y_1 and Y_0 are independent of D conditional on X, implies that

$$E[G(Y_j, y)|X = x] = E[G(Y, y)|X = x, D = j]$$

for $j = 0, 1$. Therefore, (2.5) holds under the unconfoundedness assumption and tests of (2.6) and (2.8) provide tests for (2.1) and (2.3), provided that (2.7) is satisfied.

Finally, our tests are applicable to the local average treatment effect (LATE) setup of Imbens and Angrist (1994), which is an important special case of selection on unobservables (see, e.g. Section 6 of Imbens and Wooldridge (2009)). The LATE setup presumes the existence of a binary instrument variable (IV), say Z, for the treatment assignment. Then as shown by Abadie (2002), testing (2.6) and (2.8) with D being replaced by Z provides tests for conditional treatment effects for "compliers", individuals who comply with their actual assignment of treatment and would have complied with the alternative assignment. This is because under the LATE setup, we have that

$$\begin{aligned} &E[G(Y_1, y) - G(Y_0, y)|X = x, \text{Population} = \text{Compliers}] \\ &= \frac{E[G(Y, y)|X = x, Z = 1] - E[G(Y, y)|X = x, Z = 0]}{E[D|X = x, Z = 1] - E[D|X = x, Z = 0]} \end{aligned} \tag{2.9}$$

and that the denominator on the right-hand side in (2.9) is assumed to be always strictly positive in the LATE setup.

We conclude this section by making some remarks regarding how to set up null and alternative hypotheses in applications. Needless to say, it would depend on the context of actual applications which null and alternative hypotheses should be considered. However, generally speaking, when the conditional average treatment effect, $E[Y_1 - Y_0|X = x]$, is concerned, it might be natural to consider zero conditional average treatment effect as the null hypothesis, as in Crump et al. (2008). Our framework allows an applied researcher to consider both one-sided and two-sided alternatives. When an researcher expects a particular sign of the conditional average treatment effect for some individuals *ex ante*, it would be reasonable to consider a one-sided test since the one-sided test is likely to be more powerful than the two-sided test. For example, the treatment of interest is designed to increase the outcome variable on average, then one may consider testing the null hypothesis that $E[Y_1-Y_0|X = x] = 0$ for every $x \in \mathcal{W}_x$ against the alternative hypothesis that $E[Y_1-Y_0|X = x] > 0$ for some $x \in \mathcal{W}_x$. If the conditional distributional treatment effect is considered, then one may consider the null of stochastic dominance of the treatment group over the control group against the alternative of no stochastic dominance. In addition, one may consider the null of equal distributions between the treatment and control groups against the alterative of unequal distributions.[6] Last two examples are conditional versions of null and alternative hypotheses considered in Abadie (2002).

3. Test Statistics

This section describes our test statistics for the null hypotheses (2.6) and (2.8). To include all three setups considered in Section 2 in a unifying framework, let Z denote a binary random variable that can be a treatment indicator, or in some cases, a binary instrument for treatment assignment.

Define

$$\tau_0(y, x) = E[G(Y, y)|X = x, Z = 1] - E[G(Y, y)|X = x, Z = 0]. \tag{3.1}$$

Note that the null hypotheses (2.6) and (2.8) can be equivalently stated as

$$H_0 : \tau_0(y, x) \leq 0 \text{ for each } (y, x) \in \mathcal{W}, \tag{3.2}$$

$$H_0^D : \tau_0(y, x) = 0 \text{ for each } (y, x) \in \mathcal{W} \tag{3.3}$$

[6] One potential use of testing equality of distributions is to check whether random assignment of treatment is properly done in experimental data. For example, one could test the equality of distributions of pre-intervention variables between treatment and control groups conditional on some covariates. In this case, rejecting the null of equality indicates that there is some failure to achieve the random assignment of treatment.

with the alternative hypotheses given by the negation of (3.2) and (3.3). For (3.2), we consider a class of tests based on

$$\hat{T} = \int_{\mathbb{R}^d} \int_{\mathbb{R}} \sqrt{n} \max\{\hat{\tau}(y,x), 0\}\, w(y,x)\, d\mu_y(y) d\mu_x(x), \tag{3.4}$$

where $\hat{\tau}(y,x)$ is a uniformly consistent estimator of $\tau_0(y,x)$, $w(y,x)$ is a weight function that has its support $\mathcal{W}$, and μ_y and μ_x are some measures for y and x, respectively. For (3.3), we consider a class of tests based on

$$\hat{D} = \int_{\mathbb{R}^d} \int_{\mathbb{R}} \sqrt{n}\, |\hat{\tau}(y,x)|\, w(y,x)\, d\mu_y(y) d\mu_x(x). \tag{3.5}$$

As a baseline case, we consider the case when the distributions of Y and X are absolutely continuous with respect to Lebesgue measure. In this case, we use the Lebesgue measure for μ_y and μ_x. If either the distribution of Y or the distribution of some elements of X is discrete, we can modify the integrals in the statistics $\hat{T}$ and $\hat{D}$ by using some product measure between the Lebesgue and the counting measures.[7]

To construct the test statistics $\hat{T}$ and $\hat{D}$, it is necessary to estimate $\tau_0(y,x)$. There are several alternatives to estimating $\tau_0(y,x)$. Specifically, we consider a kernel estimator of $\tau(y,x)$. That is,

$$\hat{\tau}(y,x) = \hat{E}[G(Y,y)|X=x, Z=1] - \hat{E}[G(Y,y)|X=x, Z=0],$$

where $\hat{E}[A|B]$ denote the usual kernel estimator of the conditional mean function $E[A|B]$. To describe our estimator of $\tau_0(y,x)$ in a simple form, define $p_j(x) := \Pr(Z=j|X=x)f(x)$ for $j = 0, 1$ and

$$\phi(x,z) := \frac{1(z=1)}{p_1(x)} - \frac{1(z=0)}{p_0(x)},$$

where $f(x)$ denotes the density of X. Then $\tau_0(y,x)$ is estimated by the statistic:

$$\hat{\tau}(y,x) = n^{-1} \sum_{i=1}^{n} G(Y_i, y) \hat{\phi}(x, Z_i) K_h(x - X_i),$$

where

$$\hat{\phi}(x,z) = \frac{1(z=1)}{\hat{p}_1(x)} - \frac{1(z=0)}{\hat{p}_0(x)}, \tag{3.6}$$

$$\hat{p}_j(x) = n^{-1} \sum_{i=1}^{n} 1(Z_i = j) K_h(x - X_i), \tag{3.7}$$

[7] Our test statistics are based on L_1-type functionals of nonparametric kernel estimators. Alternatively, one may consider supremum-type test statistics. It is an open question how to develop general asymptotic theory using supremum-type statistics for testing (3.2) and (3.3).

and $K_h(\cdot) = K(\cdot/h)/h^d$. Here, K is a d-dimensional kernel function, h is a bandwidth, and d is the dimension of X.

4. Asymptotic Theory

This section provides asymptotic theory for our statistics $\hat{T}$ and $\hat{D}$ both when the null hypothesis is true and when it is false. First, we show in Section 4.1 that, when suitably normalized, the statistics $\hat{T}$ and $\hat{D}$ are asymptotically distributed as the standard normal under the null hypotheses (3.2) and (3.3), respectively. Second, in Section 4.3, we show that our tests are consistent against general fixed alternatives and also show that our tests have non-trivial power against some $n^{-1/2}$ sequences of local alternatives. For notational simplicity, unless it is specified otherwise, we sometimes use the indefinite integral $\int \cdots \int$ notation to denote $\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty}$.

4.1. **Assumptions and the Asymptotic Null Distribution.**

Let

$$K_*(t) = \int K(\xi) K(\xi + t) \, d\xi \quad \text{and} \quad \rho_0(t) = \frac{K_*(t)}{K_*(0)}.$$

For y, y', x, define

$$\mu_1(y, y', x) := \sum_{j \in \{0,1\}} \frac{E[G(Y,y)G(Y,y')|X = x, Z = j]}{p_j(x)}, \tag{4.1}$$

$$\mu_2(y, y', x) := \sum_{j \in \{0,1\}} \frac{E[G(Y,y)|X = x, Z = j] \, E[G(Y,y')|X = x, Z = j]}{p_j(x)}. \tag{4.2}$$

Also, define

$$\begin{aligned}
\rho_1(y, y', x, t) &= \{\mu_1(y, y', x) - \mu_2(y, y', x)\} K_*(t), \\
\rho_2(y, x) &= \{\mu_1(y, y, x) - \mu_2(y, y, x)\} K_*(0), \\
\bar{\rho}(y, y', x) &= \frac{\{\mu_1(y, y', x) - \mu_2(y, y', x)\}}{\sqrt{\{\mu_1(y, y, x) - \mu_2(y, y, x)\}\{\mu_1(y', y', x) - \mu_2(y', y', x)\}}}, \\
\rho(y, y', x, t) &= \frac{\rho_1(y, y', x, t)}{\sqrt{\rho_2(y, x)\rho_2(y', x)}} = \bar{\rho}(y, y', x)\rho_0(t).
\end{aligned}$$

Assumption 4.1. i. *The distribution of $X \in \mathbb{R}^d$ is absolutely continuous with respect to Lebesgue measure and the probability density function f of X is continuously differentiable;*

ii. *The distribution of Y is absolutely continuous with respect to Lebesgue measure;*

iii. *$w(\cdot, \cdot)$ is a continuous function with compact support $\mathcal{W} = \mathcal{W}_y \times \mathcal{W}_x$, where $\mathcal{W}_y$ is a strict subset of $\mathcal{Y}$ and $\mathcal{W}_x$ is a strict subset of $\mathcal{X}$;*

iv. *(a)* $p_1(\cdot)$ *and* $p_0(\cdot)$ *are bounded away from zero on* $\mathcal{W}_x$ *and* $\rho_2(\cdot,\cdot)$ *is bounded away from zero on* $\mathcal{W}$*; (b)* $\bar{\rho}(y,y',x)$ *satisfies* $\bar{\rho}(y,y',x) = 1 - c_1(x)\,|y-y'|^{\alpha_1} + o(|y-y'|^{\alpha_1})$ *uniformly in* $x \in \mathcal{W}_x$ *as* $|y-y'| \to 0$ *for some positive constants* $c_1(x)$ *and* α_1 *such that* $c_1(\cdot)$ *is bounded away from 0 on* $\mathcal{W}_x$*.*

v. *(a)* K *is a s-order kernel function with support* $\{u \in \mathbb{R}^d : \|u\| \le 1/2\}$*, symmetric around zero, integrates to 1 and is s-times continuously differentiable, where s is an integer that satisfies* $s > 3d/2$*; (b) The kernel satisfies* $\rho_0(t) = 1 - c_0\,\|t\|^{\alpha_0} + o(\|t\|^{\alpha_0})$ *as* $t \to 0$ *for some positive constants* c_0 *and* α_0*.*

vi. *As functions of* x*,* $E[G(Y,y)|X=x,Z=j]$*,* $f(x)$*,* $p_j(x)$ *for* $j = 0,1$ *are s-times continuously differentiable for each* y *with uniformly bounded derivatives;*

vii. $\sup_{(y,x)\in\mathcal{W}} E\left[|G(Y,y)|^3\,|X=x,Z=j\right] < \infty$ *for* $j = 0,1$*;*

viii. $\{G(\cdot,y) : y \in \mathcal{W}_y\}$ *is a VC class of functions with an envelope function* $\mathbf{M}$ *satisfying* $\sup_{x\in\mathcal{W}_x} E[\mathbf{M}^2(Y)|X=x] < \infty$*;*

ix. *The bandwidth satisfies* $nh^{2s} \to 0$*,* $nh^{3d} \to \infty$ *and* $\left(nh^{2d}\right)^{1/2}/\log n \to \infty$*, where* $s > 3d/2$*.*

We make some comments regarding the regularity conditions. Most of them are standard in the literature on kernel estimation. Conditions (i) and (ii) are just convenient assumptions to present our main result. It is straightforward to extend them to more general settings. For example, if the distribution of Y or the distribution of some elements of X (but not all) is discrete, we can modify the test statistics $\hat{T}$ and $\hat{D}$ defined in (3.4) and (3.5) with counting measure in proper directions in a straightforward way.

Condition (iii) assumes continuity of the weight function and also assumes that $\mathcal{W}$ is a strict compact subset of the support of (Y,X). Given the latter condition, it is reasonable to assume Condition (iv). Part (b) of condition (iv) is automatically satisfied if $G(Y,y) \equiv -Y$. The compact support assumption on $\mathcal{W}$ is needed to carry out studentization of test statistics. This is a stringent assumption that keeps us from testing the null hypothesis on the entire support. However, in practice, it would be difficult to estimate $\hat{\tau}(y,x)$ with good precision at boundary points, and therefore, there is not much of loss of generality by assuming that $\mathcal{W}$ is compact.

Condition (v) can be satisfied easily by choosing a suitable kernel function, and conditions (vi) and (vii) impose some smoothness assumptions and moment restrictions on the underlying true data generating process. Condition (viii) is satisfied if $G(Y,y) \equiv -Y$ or if $G(Y,y) = 1(Y \le y)$. In view of condition (ix), when $d = 1$, a usual second-order kernel can be used with the bandwidth condition such that $h = c_1 n^{-\delta}$ for some positive constant c_1 with $1/4 < \delta < 1/3$.[8]

We first show that, in the least favorable case of the null hypothesis (3.2) (i.e., the case where $\tau_0(y,x) = 0$ for each $(y,x) \in \mathcal{W}$), the asymptotic distribution of $\hat{T}$ is the standard normal in the sense that

$$\frac{\hat{T} - a_n}{\sigma_0} \xrightarrow{d} N(0,1),$$

with the asymptotic bias and variance of $\hat{T}$ given by, respectively,

$$a_n := h^{-d/2} \int_{\mathbb{R}^d} \int_{\mathbb{R}} \sqrt{\rho_2(y,x)} w(y,x) dy dx \cdot E \max\{\mathbb{Z}_1, 0\}, \tag{4.3}$$

$$\sigma_0^2 := \int_{T_0} \int_{\mathbb{R}^d} \int_{\mathbb{R}} \int_{\mathbb{R}} F[\rho(y,y',x,t)] \sqrt{\rho_2(y,x)\rho_2(y',x)} w(y,x) w(y',x) dy dy' dx dt, \tag{4.4}$$

$$F(\rho) := Cov\left(\max\{\sqrt{1-\rho}\mathbb{Z}_1 + \rho\mathbb{Z}_2, 0\}, \max\{\mathbb{Z}_2, 0\}\right). \tag{4.5}$$

Here, $\mathbb{Z}_1$ and $\mathbb{Z}_2$ denote mutually independent standard normal random variables and $T_0 := \{t \in \mathbb{R}^d : \|t\| \leq 1\}$. For practical implementation of our test, we need to estimate the asymptotic bias and variance consistently. First of all, by calculus, $E \max\{\mathbb{Z}_1, 0\} = 1/\sqrt{2\pi} \approx 0.39894$. Note that $F(\rho)$ that appears in the definition of σ_0^2 can be approximated for each value of ρ with arbitrary accuracy by simulating a large number of independent standard normal random variables $(\mathbb{Z}_1, \mathbb{Z}_2)$. On the other hand, under the null hypothesis (3.3), the asymptotic distribution of $\hat{D}$ is given by

$$\frac{\hat{D} - a_{n,D}}{\sigma_{0,D}} \xrightarrow{d} N(0,1),$$

with

$$a_{n,D} = h^{-d/2} \int_{\mathbb{R}^d} \int_{\mathbb{R}} \sqrt{\rho_2(y,x)} w(y,x) dy dx \cdot E|\mathbb{Z}_1|, \tag{4.6}$$

$$\begin{aligned} \sigma_{0,D}^2 = & \int_{T_0} \int_{\mathbb{R}^d} \int_{\mathbb{R}} \int_{\mathbb{R}} Cov\left(\left|\sqrt{1-\rho^2(y,y',x,t)}\mathbb{Z}_1 + \rho(y,y',x,t)\mathbb{Z}_2\right|, |\mathbb{Z}_2|\right) \\ & \times \sqrt{\rho_2(y,x)\rho_2(y',x)} w(y,x) w(y',x) dy dy' dx dt, \end{aligned} \tag{4.7}$$

where $\mathbb{Z}_1, \mathbb{Z}_2$ and $T_0 = \{t \in \mathbb{R}^d : \|t\| \leq 1\}$ are defined as above. By calculus, $E|\mathbb{Z}_1| = 2/\sqrt{2\pi} \approx 0.79788$.

[8] Methods for selecting h in applications are not yet available. We provide some simulation evidence regarding sensitivity to the choice of h in Section 6. Generally speaking, an optimal bandwidth for nonparametric testing is different from one for nonparametric estimation. For example, in order to capture tradeoffs between the size and power, Gao and Gijbels (2008) derive a bandwidth-selection rule by utilizing an Edgeworth expansion of the asymptotic distribution of the test statistic concerned. The results of Gao and Gijbels (2008) are not directly applicable to our tests.

In the case that $G(Y,y) = -Y$, we have

$$\rho(y,y',x,t) = \rho_0(t),$$

so that the expressions for a_n and σ_0^2 are simplified to:

$$\begin{aligned} a_n &= h^{-d/2} \int_{\mathbb{R}^d} \sqrt{\rho_2(x)} w(x) dx \cdot \frac{1}{\sqrt{2\pi}}, \\ \sigma_0^2 &= \int_{T_0} F\left[\rho_0(t)\right] dt \int_{\mathbb{R}^d} \rho_2(x) w^2(x) dx, \end{aligned} \tag{4.8}$$

where $w(x)$ is a weight function for x and

$$\rho_2(x) = K_*(0) \sum_{j \in \{0,1\}} \frac{E\left[Y^2 | X = x, Z = j\right] - \left(E\left[Y | X = x, Z = j\right]\right)^2}{p_j(x)}.$$

Likewise, analogous simplification occurs for the two-sided test.

The unknown quantities $\rho_1(y,y',x,t), \rho_2(y,x)$ and $\rho(y,y',x,t)$ that appear in (4.3) - (4.7) can be estimated nonparametrically by:

$$\begin{aligned} \hat{\rho}_1(y,y',x,t) &= \left\{\hat{r}_1(y,y',x) - \hat{r}_2(y,y',x)\right\} K_*(t), \\ \hat{\rho}_2(y,x) &= \left\{\hat{r}_1(y,y,x) - \hat{r}_2(y,y,x)\right\} K_*(0), \\ \hat{\rho}(y,y',x,t) &= \frac{\hat{\rho}_1(y,y',x,t)}{\sqrt{\hat{\rho}_2(y,x)\hat{\rho}_2(y',x)}}, \end{aligned}$$

where

$$\begin{aligned} \hat{r}_1(y_1,y_2,x) &= \sum_{j \in \{0,1\}} \sum_{i=1}^n \frac{G(Y_i,y_1) G(Y_i,y_2) 1(Z_i = j) K_h(x - X_i)}{n \hat{p}_j^2(x)}, \\ \hat{r}_2(y_1,y_2,x) &= \sum_{j \in \{0,1\}} \sum_{i=1}^n \sum_{k=1}^n \frac{G(Y_i,y_1) G(Y_k,y_2) 1(Z_i = j) 1(Z_k = j) K_h(x - X_i) K_h(x - X_k)}{n^2 \hat{p}_j^3(x)}, \end{aligned}$$

and $\hat{p}_j(x)$ is defined as in (3.7). With these definitions, we estimate (a_n, σ_0^2) and $(a_{n,D}, \sigma_{0,D}^2)$ by:

$$\begin{aligned} \hat{a}_n &= h^{-d/2} \int_{\mathbb{R}^d} \int_{\mathbb{R}} \sqrt{\hat{\rho}_2(y,x)} w(y,x) dy dx \cdot E \max\{\mathbb{Z}_1, 0\}, \\ \hat{\sigma}^2 &= \int_{T_0} \int_{\mathbb{R}^d} \int_{\mathbb{R}} \int_{\mathbb{R}} Cov\left(\max\left\{\sqrt{1 - \hat{\rho}^2(y,y',x,t)}\mathbb{Z}_1 + \hat{\rho}(y,y',x,t)\mathbb{Z}_2, 0\right\},\ \max\{\mathbb{Z}_2, 0\}\right) \\ &\quad \times \sqrt{\hat{\rho}_2(y,x)\hat{\rho}_2(y',x)} w(y,x) w(y',x) dy dy' dx dt. \end{aligned}$$

and

$$\hat{a}_{n,D} = h^{-d/2} \int_{\mathbb{R}^d} \int_{\mathbb{R}} \sqrt{\hat{\rho}_2(y,x)} w(y,x) dy dx \cdot E\left|\mathbb{Z}_1\right|,$$

$$\hat{\sigma}_D^2 = \int_{T_0}\int_{\mathbb{R}^d}\int_{\mathbb{R}}\int_{\mathbb{R}} Cov\left(\left|\sqrt{1-\hat{\rho}^2(y,y',x,t)}\mathbb{Z}_1+\hat{\rho}(y,y',x,t)\mathbb{Z}_2\right|,\ |\mathbb{Z}_2|\right)$$
$$\times\sqrt{\hat{\rho}_2(y,x)\hat{\rho}_2(y',x)}w(y,x)w(y',x)dydy'dxdt,$$

respectively. The integrals appearing above and in the definitions of the statistics $\hat{T}$ and $\hat{D}$ can be evaluated using the composite trapezoid rule or more sophisticated numerical methods such as Monte Carlo simulation especially when the dimension of X is high.

Now, we consider standardized test statistics of the form

$$\hat{S} = \frac{\hat{T}-\hat{a}_n}{\hat{\sigma}} \text{ and } \hat{S}_D = \frac{\hat{D}-\hat{a}_{n,D}}{\hat{\sigma}_D}. \tag{4.9}$$

Our tests are based on the following decision rules:

$$\text{Reject } H_0 \text{ if } \hat{S} > z_{1-\alpha}\ ,$$
$$\text{Reject } H_0^D \text{ if } \hat{S}_D > z_{1-\alpha}$$

at the nominal significance level α, where z_α is the α quantile of the standard normal distribution for $0<\alpha<1$. The following theorem shows that our tests have an asymptotically valid size:

Theorem 4.1. *Let Assumption 4.1 hold. Then, (a) under the null hypothesis H_0,*

$$\lim_{n\to\infty}\Pr\left(\hat{S}>z_{1-\alpha}\right)\le\alpha,$$

with equality when $\tau_0(y,x)=0$ for each $(y,x)\in\mathcal{W}$ and (b) under the null hypothesis H_0^D,

$$\lim_{n\to\infty}\Pr\left(\hat{S}_D>z_{1-\alpha}\right)=\alpha.$$

We prove Theorem 4.1 in three steps whose details are provided in Appendix:

Step 1. The asymptotic approximation of $\hat{T}$ by T_n using the uniform approximation of $\hat{\tau}(y,x)$ up to stochastic order $o_p(n^{-1/2})$, where

$$T_n := \int\int\sqrt{n}\max\{[\tau_n(y,x)-E\tau_n(y,x)],0\}w(y,x)dydx,$$
$$\tau_n(y,x) := \frac{1}{n}\sum_{i=1}^{n}\left[\{G(Y_i,y)-E[G(Y,y)|X=x,Z=1]\}\frac{1(Z_i=1)}{p_1(x)}\right.$$
$$\left.-\{G(Y_i,y)-E[G(Y,y)|X=x,Z=0]\}\frac{1(Z_i=0)}{p_0(x)}\right]K_h(x-X_i). \tag{4.10}$$

Step 2. To obtain the asymptotic distribution of $T_n^P(B)$, a *Poissonized* version T_n, where the sample size n is replaced by a Poisson random variable N with mean n that is independent of the original sequence $\{(Y_i,X_i):i\ge 1\}$ and the integral is taken over a subset B of $\mathcal{W}$.

Step 3. To *de-Poissonize* $T_n^P(B)$ to derive the asymptotic normality of $T_n(B)$, and hence that of $\hat{T}$ by letting B increase.

Steps 2-3 ("Poissonization" and "de-Poissonization") require lengthy, nontrivial derivation using the "Poissonization" technique developed in Giné et al. (2003). Although the above steps closely follow those of Giné et al. (2003), we need to extend their results to our testing problem with general multi-dimensional variates $d \geq 1$ and statistics that are different from the L_1- norm.[9] See Anderson et al. (2009) and Mason and Polonik (2009) for different applications of the "Poissonization" technique.

4.2. **The Weight Function.** In this section, we consider the choice of the weight function $w(y, x)$. There could be potentially many functions one could consider, but at least the following three functions seem to be natural:

(1) $w_1(y, x) = 1$ on $\mathcal{W}$ (a uniform weight function),
(2) $w_2(y, x) = [\rho_2(y, x)]^{-1/2}$ (an inverse-variance weight function),
(3) $w_3(y, x) = p_1(x) \cdot p_0(x)$ (a density weight function).

The uniform weight function is simple and there is no need to estimate the unknown population components. The inverse-variance weight function is a reasonable candidate as a weight function since it weighs down the values of (y, x) for which $\tau(y, x)$ is estimated imprecisely. As it can be seen from (4.8), the asymptotic distribution of the test statistics with $G(Y, y) \equiv -Y$ would be completely free from nuisance parameters if $w(x) = [\rho_2(x)]^{-1/2}$. The density weight function $p_1(x) \cdot p_0(x)$ is a convenient choice that would remove the random denominators in $\hat{\tau}(y, x)$. Therefore, in this case, it might be possible to take $\mathcal{W}$ to be the whole support of (Y, X) (i.e. $\mathcal{W} \equiv \mathcal{Y} \times \mathcal{X}$); however, details are not worked out in the paper. The asymptotic theory developed in Section (4.1) assumes that $w(y, x)$ is known. It is straightforward to show that the asymptotic null distribution is the same with an estimated $w(y, x)$ if an estimator of $w(y, x)$ is uniformly consistent at a uniform rate of $o_p(h^{d/2})$ and $w(y, x)$ is bounded from above and below from zero on $\mathcal{W}$.

4.3. **The Asymptotic Power Properties.** In this section, we investigate power properties of our tests. We first establish that the tests $\hat{S}_D$ and $\hat{S}_D$ are consistent against the fixed alternative hypotheses

$$H_1 \quad : \quad \int\int \max\{\tau_0(y, x), 0\} \, w(y, x) dy dx > 0, \tag{4.11}$$

$$H_1^D \quad : \quad \int\int |\tau_0(y, x)| \, w(y, x) dy dx > 0, \tag{4.12}$$

[9] We have considered L_1-type functionals to construct our test statistics. More generally, one may consider L_p-type functionals with $p \geq 1$. The corresponding asymptotic theory would be different from that obtained in this paper.

respectively, where τ_0 is defined in (3.1).

Theorem 4.2. *Let Assumption 4.1 hold. Then, (a) under the alternative hypothesis H_1,*

$$\lim_{n\to\infty} \Pr\left(\hat{S} > z_{1-\alpha}\right) = 1$$

and (b) under the alternative hypothesis H_1^D,

$$\lim_{n\to\infty} \Pr\left(\hat{S}_D > z_{1-\alpha}\right) = 1.$$

Next, we determine the power of $\hat{S}$ and $\hat{S}_D$ against some sequences of local alternatives. Consider the following sequences of local alternatives converging to the null hypothesis at the rate $n^{-1/2}$:

$$H_a : \tau_0(y,x) = n^{-1/2}\delta(y,x), \tag{4.13}$$

$$H_a^D : \tau_0(y,x) = n^{-1/2}\delta_D(y,x), \tag{4.14}$$

where $\delta(\cdot,\cdot)$ is a real non-negative function satisfying $\int\int \delta(y,x)w(y,x)dydx > 0$ and $\delta_D(\cdot,\cdot)$ is a real function satisfying $\int\int |\delta_D(y,x)|\, w(y,x)dydx > 0$. Under H_a, we show that

$$\frac{\hat{T} - \tilde{a}_n}{\sigma_0} \xrightarrow{d} N(0,1),$$

where σ_0 is defined as in (4.4) and

$$\tilde{a}_n = \int\int E\max\left\{\delta(y,x) + h^{-d/2}\sqrt{\rho_2(y,x)}\mathbb{Z}, 0\right\} w(y,x)dydx.$$

See proof of Theorem 4.3 for details. Since we have

$$\hat{S} = \frac{\hat{T} - \tilde{a}_n}{\sigma_0} + \frac{\tilde{a}_n - a_n}{\sigma_0} + o_p(1)$$

under H_a and

$$\begin{aligned} &\tilde{a}_n - a_n \\ &= \int\int E\left[\max\left\{\delta(y,x) + h^{-d/2}\sqrt{\rho_2(y,x)}\mathbb{Z}, 0\right\} - \max\left\{h^{-d/2}\sqrt{\rho_2(y,x)}\mathbb{Z}, 0\right\}\right] w(y,x)dydx \\ &\geq \frac{1}{2}\int\int \delta(y,x)w(y,x)dydx > 0, \end{aligned}$$

we expect that our test $\hat{S}$ is powerful against H_a. Similarly, under H_a^D, we can show that

$$\frac{\hat{D} - \tilde{a}_{n,D}}{\sigma_{0,D}} \xrightarrow{d} N(0,1),$$

where $\sigma_{0,D}$ is defined as in (4.7) and

$$\tilde{a}_{n,D} = \int\int E\left|\delta(y,x) + h^{-d/2}\sqrt{\rho_2(y,x)}\mathbb{Z}\right| w(y,x)dydx.$$

Since we have

$$\begin{aligned}\tilde{a}_{n,D} - a_{n,D} &= \int\int E\left[\left|\delta(y,x) + h^{-d/2}\sqrt{\rho_2(y,x)}\mathbb{Z}\right| - h^{-d/2}\sqrt{\rho_2(y,x)}\left|\mathbb{Z}\right|\right] w(y,x)dydx \\ &\geq 0\end{aligned}$$

by the Anderson's lemma (see, e.g., van der Vaart and Wellner (1996, Lemma 3.11.4)), we also expect that the test $\hat{S}_D$ is powerful against H_a^D.

The following theorem formally establishes that our tests have non-trivial local power against H_a and H_a^D in the sense that they are asymptotically locally unbiased.

Theorem 4.3. *Let Assumption 4.1 hold. Then, (a) under the alternative hypothesis H_a ,*

$$\lim_{n\to\infty} \Pr\left(\hat{S} > z_{1-\alpha}\right) > \alpha$$

and (b) under the alternative hypothesis H_a^D,

$$\lim_{n\to\infty} \Pr\left(\hat{S}_D > z_{1-\alpha}\right) \geq \alpha.$$

5. Extensions

In Section 5.1, we show that, in the case when the null hypothesis H_0 is expressed in terms of inequality constraints as in (3.2), we can develop a test that is (locally) more powerful than the test $\hat{S}$ defined in (4.9). In Section 5.2, we describe an extension of our tests for quantile treatment effects.

5.1. A more powerful test for the null hypothesis with inequality constraints.

Define

$$C := \{(y,x) \in \mathcal{W} : \tau_0(y,x) = 0\} \tag{5.1}$$

to be the subset of $\mathcal{Y} \times \mathcal{X}$ on which the null hypothesis (2.6) holds with equality. Note that C can be written as

$$C := \{(y,x) \in C_1(x) \times \mathcal{W}_x\},$$

where $C_1(x) := \{y \in \mathcal{W}_y : \tau_0(y,x) = 0\}$ for each $x \in \mathcal{W}_x$.

It turns out that the asymptotic distribution of $\hat{T}$ depends on C when $\int\int_C w(y,x)dydx > 0$. In this case, we can show that the asymptotic bias and variance of $\hat{T}$ (defined in (3.4)) are given by

(5.2)

$$a_n(C) = h^{-d/2} \int_{\mathbb{R}^d} \int_{C_1(x)} \sqrt{\rho_2(y,x)} w(y,x) dy dx \cdot E \max\{\mathbb{Z}_1, 0\},$$

(5.3)

$$\sigma_0^2(C) = \int_{T_0} \int_{\mathbb{R}^d} \int_{C_1(x)} \int_{C_1(x)} F[\rho(y,y',x,t)] \sqrt{\rho_2(y,x)\rho_2(y',x)} w(y,x) w(y',x) dy dy' dx dt,$$

respectively, where $\mathbb{Z}_1$ and $\mathbb{Z}_2$ are mutually independent standard normal random variables and $T_0 = \{t \in \mathbb{R}^d : \|t\| \le 1\}$ as in (4.3) and (4.4). This implies that, when C is a non-negligible set, we may construct a less conservative test than $\hat{S}$ using the bias and variance formulae of (5.2) and (5.3).

In general, the set C is unknown and has to be estimated. As we explain below, it is difficult to estimate C. In this paper, we estimate an outer set C_ϵ of C, that is $C_\epsilon := \{(y,x) \in \mathcal{W} : |\tau_0(y,x)| \le \epsilon\}$ for some small constant $\epsilon > 0$. To be precise, we define the estimator of C_ϵ to be

$$\hat{C}_\epsilon := \{(y,x) \in \hat{C}_{1\epsilon}(x) \times \mathcal{W}_x\},$$

where $\hat{C}_{1\epsilon}(x) := \{y \in \mathcal{W}_y : |\hat{\tau}(y,x)| \le \eta_n + \epsilon\}$ for each $x \in \mathcal{W}_x$. Here, η_n is a sequence of positive constants that converges to zero at a rate satisfying Assumption 5.1 below. When $\int\int_{\hat{C}_\epsilon} w(y,x) dy dx > 0$, we can estimate $a_n(C_\epsilon)$ and $\sigma_0^2(C_\epsilon)$ by

$$\begin{aligned}
\hat{a}_n(\hat{C}_\epsilon) &= h^{-d/2} \int_{\mathbb{R}^d} \int_{\hat{C}_{1\epsilon}(x)} \sqrt{\hat{\rho}_2(y,x)} w(y,x) dy dx \cdot E \max\{\mathbb{Z}_1, 0\}, \\
\hat{\sigma}^2(\hat{C}_\epsilon) &= \int_{T_0} \int_{\mathbb{R}^d} \int_{\hat{C}_{1\epsilon}(x)} \int_{\hat{C}_{1\epsilon}(x)} Cov\left(\max\{\sqrt{1-\hat{\rho}^2(y,y',x,t)}\mathbb{Z}_1 + \hat{\rho}(y,y',x,t)\mathbb{Z}_2, 0\}, \max\{\mathbb{Z}_2, 0\} \right) \\
&\quad \times \sqrt{\hat{\rho}_2(y,x)\hat{\rho}_2(y,x)} w(y,x) w(y',x) dy dy' dx dt,
\end{aligned}$$

where $\hat{\rho}_2(y,x)$ and $\hat{\rho}^2(y,y',x,t)$ are defined as in Section 3.1. In this case, we let

$$\hat{S}_C = \frac{\hat{T} - \hat{a}_n(\hat{C}_\epsilon)}{\hat{\sigma}(\hat{C}_\epsilon)}.$$

Notice that, when $\int\int_{\hat{C}_\epsilon} w(y,x) dy dx = 0$, the estimators $\hat{a}_n(\hat{C}_\epsilon)$ and $\hat{\sigma}^2(\hat{C}_\epsilon)$ are degenerate at zero and hence $\hat{S}_C$ is not well-defined. However, the test $\hat{S}$ based on the least favorable case is always well-defined and has an asymptotically valid size, though it may be conservative when C_ϵ is a strict subset of $\mathcal{W}$. Therefore, we suggest the following decision rule:

$$\text{Reject } H_0 \text{ if } \hat{S}^* > z_{1-\alpha},$$

where

$$\hat{S}^{*}=\begin{cases}\hat{S}_{C} & \text{if } \int\int_{\hat{C}_{\epsilon}} w(y,x)dydx>0\\ \hat{S} & \text{if } \int\int_{\hat{C}_{\epsilon}} w(y,x)dydx=0\end{cases}.$$

To investigate the size and power performance of $\hat{S}^{*}$, in addition to Assumption 4.1, we need to impose the following regularity conditions on the contact set C and the tuning parameter η_n.

Assumption 5.1. i. *Whenever the Lebesgue measure $\lambda(C_\epsilon)$ of C_ϵ is strictly positive, the boundary of C_ϵ satisfies $h^*(t) := \lambda(\{(y,x) : \epsilon < |\tau_0(y,x)| \le \epsilon + t\}) = O(t^\gamma)$ as $t \to 0$ for some constants $\epsilon > 0$ and $\gamma > 0$.*

ii. *The tuning parameter η_n satisfies $nh^d\eta_n^{2+2\gamma}/(\log n)^2 \to 0$ and $nh^{2d}\eta_n^2/(\log n)^2 \to \infty$ as $n \to \infty$.*

To appreciate the degrees of restrictions behind Assumption 5.1 (i), consider the following example of $\tau_0(y,x)$ that satisfies Assumption 5.1 (i):

$$\tau_0(y,x)=\begin{cases}-x^{1/\gamma_0} & \text{if } x>0\\ 0 & \text{if } x\le 0\end{cases},$$

where $x \in \mathcal{W}_x \equiv [-R,R] \subset \mathbb{R}$ with some constant $R > 0$. In this example, the larger γ_0 is, the less smooth τ_0 is around the "contact point" at zero ($x = 0$). Note that roughly speaking, in this example, $s < 1/\gamma_0$. Recall that in Assumption 4.1 (ix), we need to assume that $s > 3d/2$. As d gets large, we need larger s and also larger γ to satisfy both Assumptions 4.1 (ix) and 5.1 (ii). Thus, this is impossible if we consider $\epsilon = 0$, i.e. estimation of C rather than C_ϵ with $\epsilon > 0$. For the latter, regardless of the smoothness of τ_0 around zero, one can choose $\gamma = 1$. This is the reason why we consider estimation of the outer set C_ϵ of the contact set C. A similar difficulty arises in nonparametric estimation of an argmin set in the partial identification setup (Chernozhukov et al., 2009).

Suppose that $h_n \propto n^{-\delta}$ for some constant $\delta > 0$. Then Assumption 5.1 (ii) is satisfied, for example, if $\eta_n \propto n^{-\nu}$ with

$$\frac{1-d\delta}{2+2\gamma}<\nu<\frac{1-2d\delta}{2},$$

provided that $\gamma > (d\delta)/(1-2d\delta)$. The constant term $(1-2d\delta)$ has to be positive under Assumption 4.1 (ix). Thus, Assumption 5.1 (ii) is less stringent with a larger γ. For example, if $\gamma = 1$, it requires that $\delta < 1/(3d)$.

The following theorem shows that the test $\hat{S}^{*}$ has an asymptotically valid size under the null hypothesis (3.2) and consistent against the fixed alternative (4.11).

Theorem 5.1. *Suppose that Assumptions 4.1 and 5.1 hold. Then, (a) under the null hypothesis (3.2),*

$$\lim_{n\to\infty} \Pr\left(\hat{S}^* > z_{1-\alpha}\right) \le \alpha,$$

and (b) under the alternative hypothesis (4.11),

$$\lim_{n\to\infty} \Pr\left(\hat{S}^* > z_{1-\alpha}\right) = 1.$$

To investigate the local power properties of $\hat{S}^*$, we consider the following sequence of local alternatives:

$$H_a^* : \tau_0(y,x) = \mu(y,x) + n^{-1/2}\delta(y,x). \tag{5.4}$$

Assumption 5.2. *The functions $\mu(\cdot,\cdot)$ and $\delta(\cdot,\cdot)$ satisfy the following:*

i. $\int\int_{C_a} w(y,x)dydx > 0$, *where* $C_a = \{(y,x) \in \mathcal{W} : -\epsilon \le \mu(y,x) \le 0\}$ *and* $\epsilon > 0$ *is the same constant as in Assumption 5.1.*
ii. $\sup_{(y,x)\in\mathcal{W}} \mu(y,x) \le 0$.
iii. $\delta(\cdot,\cdot)$ *is a non-negative function with*

$$\int\int_{C_a} \delta(y,x)w(y,x)dydx > 0 \text{ and } \sup_{(y,x)\in\mathcal{W}} \delta(y,x) < \infty.$$

iv. *The boundary of C_a satisfies* $h^{**}(t) := \lambda(\{(y,x) : \epsilon < |\mu(y,x)| \le t+\epsilon\}) = O(t^\gamma)$ *as* $t \to 0$ *for some constant* $\gamma > 0$.

The local alternative hypothesis H_a^* in (5.4) is more general than the hypothesis H_a in (4.13) in the sense that H_a^* allows $\mu(y,x)$ to be strictly negative for some $(y,x) \in \mathcal{W}$, whereas H_a sets $\mu(y,x) = 0$ for each (y,x). The following theorem shows that, under the sequence of local alternatives H_a^*, the test $\hat{S}^*$ is strictly unbiased and can be more powerful than $\hat{S}$ when C_a is a strict subset of $\mathcal{W}$.

Theorem 5.2. *Suppose that Assumptions 4.1, 5.1 (ii) and 5.2 hold. Then, under the alternative hypothesis H_a^*, we have (a)*

$$\lim_{n\to\infty} \Pr\left(\hat{S}^* > z_{1-\alpha}\right) > \alpha$$

and (b)

$$\lim_{n\to\infty} \Pr\left(\hat{S}^* > z_{1-\alpha}\right) > \lim_{n\to\infty} \Pr\left(\hat{S} > z_{1-\alpha}\right), \tag{5.5}$$

provided

$$\int\int \sqrt{\rho_2(y,x)}w(y,x)dydx > \int\int_{C_a} \sqrt{\rho_2(y,x)}w(y,x)dydx.$$

5.2. **Testing for quantile treatment effects.** Quantile treatment effects are increasingly popular in empirical research. A recent literature includes, for example, Abadie et al. (2002); Chernozhukov and Hansen (2005); Firpo (2007); Chernozhukov et al. (2009), among others.

To develop tests for quantile treatment effects, let $Q_0(\tau|x,z)$ denote the τ-th quantile of Y conditional on $X = x$ and $Z = z$ for $\tau \in (0,1)$. Let

$$\theta_0(\tau, x) = Q_0(\tau|x, 0) - Q_0(\tau|x, 1). \tag{5.6}$$

Then quantile analogs of the null hypotheses (3.2) and (3.3) are:

$$H_{0q} : \theta_0(\tau, x) \leq 0 \text{ for each } (\tau, y) \in \mathcal{T} \times \mathcal{W}_x, \tag{5.7}$$

$$H_{0q}^D : \theta_0(\tau, x) = 0 \text{ for each } (\tau, y) \in \mathcal{T} \times \mathcal{W}_x, \tag{5.8}$$

where $\mathcal{T}$ is a strict compact subset of $(0,1)$. As in Section 3, for (5.7) we consider a class of tests based on

$$\hat{T}_q = \int_{\mathbb{R}^d} \int_{\mathbb{R}} \sqrt{n} \max \left\{ \hat{\theta}(\tau, x), 0 \right\} w_q(\tau, x)\, d\tau d\mu_x(x), \tag{5.9}$$

where $\hat{\theta}(\tau, x)$ is a uniformly consistent estimator of $\theta_0(\theta, x)$, $w_q(\theta, x)$ is a weight function that has its support $\mathcal{T} \times \mathcal{W}_x$. For (5.8), we consider a class of tests based on

$$\hat{D}_q = \int_{\mathbb{R}^d} \int_{\mathbb{R}} \sqrt{n} \left| \hat{\theta}(\tau, x) \right| w_q(\tau, x)\, d\tau d\mu_x(x). \tag{5.10}$$

In this section, we consider the case when quantile treatment effects are evaluated under randomized controlled experiments or under the unconfoundedness assumption. That is, Y_1 and Y_0 are independent of D conditional on X, so that $Z \equiv D$ in this section. Also, we assume that the distributions of Y and X are absolutely continuous with respect to Lebesgue measure.

Let $f_{Y|X,Z}(\cdot|x,z)$ denote the probability density function of Y conditional on $X = x$ and $Z = z$. For (τ, τ', x), define

$$\begin{aligned}
\rho_{1q}(\tau, \tau', x, t) &= \left\{ \sum_{j \in \{0,1\}} \frac{\min\{\tau, \tau'\} - \tau\tau'}{f_{Y|X,Z}[Q_0(\tau|x,j)|x,j] f_{Y|X,Z}[Q_0(\tau'|x,j)|x,j] p_j(x)} \right\} K_*(t), \\
\rho_{2q}(\tau, x) &= \left\{ \sum_{j \in \{0,1\}} \frac{\tau(1-\tau)}{f_{Y|X,Z}^2[Q_0(\tau|x,j)|x,j] p_j(x)} \right\} K_*(0), \\
\rho_q(\tau, \tau', x, t) &= \frac{\rho_{1q}(\tau, \tau', x, t)}{\sqrt{\rho_{2q}(\tau, x) \rho_{q2}(\tau', x)}}, \\
\overline{\rho}_q(\tau, \tau', x) &= \frac{\rho_q(\tau, \tau', x, t)}{\rho_0(t)}.
\end{aligned}$$

In addition, define

$$a_{nq} := h^{-d/2} \int_{\mathbb{R}^d} \int_{\mathbb{R}} \sqrt{\rho_{2q}(\tau, x)} w_q(\tau, x) d\tau dx \cdot E \max \{\mathbb{Z}_1, 0\}, \tag{5.11}$$

$$\sigma_{0q}^2 := \int_{T_0} \int_{\mathbb{R}^d} \int_{\mathbb{R}} \int_{\mathbb{R}} F[\rho_q(\tau, \tau', x, t)] \sqrt{\rho_{2q}(\tau, x)\rho_{2q}(\tau', x)} w_q(\tau, x) w_q(\tau', x) d\tau d\tau' dx dt. \tag{5.12}$$

We make the following assumption.

Assumption 5.3. *Let conditions i, ii, v, ix of Assumption 4.1 hold. In addition, assume that conditions iii and iv hold with $w_q(\cdot,\cdot)$ and $\overline{\rho}_q(\tau, \tau', x)$. Suppose that a nonparametric estimator of $Q_0(\tau|x, z)$ has a Bahadur-type linear expansion of the following form:*

$$\begin{aligned} &\hat{Q}(\tau|x, j) - Q_0(\tau|x, j) \\ &= (nh^d)^{-1} \sum_{i=1}^{n} \frac{[\tau - 1\{Y_i \leq Q_0(\tau|X_i, Z_i)\}]}{f_{Y|X,Z}[Q_0(\tau|x, j)|x, j] p_j(x)} 1(Z_i = j) K\left(\frac{x - X_i}{h}\right) + R_{nj}(\tau, x), \end{aligned} \tag{5.13}$$

where for each $j = 0, 1$, $f_{Y|X,Z}[Q_0(\tau|x, j)|x, j]$ is bounded away from zero on $\mathcal{T} \times \mathcal{W}_x$ and the remainder term $R_{nj}(\tau, x)$ is of order $o_p(n^{-1/2})$ uniformly over τ and x in $\mathcal{T} \times \mathcal{W}_x$.

It is a high-level assumption to impose a Bahadur-type linear expansion for the nonparametric estimator; however, related low-level conditions can be found in the literature. See, e.g., Chaudhuri (1991), Fan et al. (1994), and Chaudhuri et al. (1997) for Bahadur-type expansions with a fixed quantile τ. As demonstrated in Hoderlein and Mammen (2009, Appendix), it is possible to make the Bahadur-type expansion uniform over τ in a compact subset of $(0, 1)$.

Theorem 5.3. *Let Assumption 5.3 hold. Then, under the null hypothesis H_{0q},*

$$\lim_{n \to \infty} \Pr\left(\frac{\hat{T}_q - a_{nq}}{\sigma_{0q}} > z_{1-\alpha}\right) \leq \alpha,$$

with equality when $\theta_0(\tau, x) = 0$ for each $(\tau, x) \in \mathcal{T} \times \mathcal{W}_x$. Furthermore, under the alternative hypothesis H_{1q}:

$$H_{1q} : \int \int \max\{\theta_0(\tau, x), 0\} w_q(\tau, x) d\tau dx > 0,$$

we have that

$$\lim_{n \to \infty} \Pr\left(\frac{\hat{T}_q - a_{nq}}{\sigma_{0q}} > z_{1-\alpha}\right) = 1.$$

This theorem establishes analogs of Theorems 4.1 and 4.2 for conditional quantile treatment effects with the statistic $\hat{T}_q$. It is rather straightforward to construct consistent estimators of a_{nq} and σ_{0q}^2 and show that a feasible version of the test has a valid size under H_{0q} and consistent under the fixed alternative. Similar results can be obtained for the test statistic $\hat{D}_q$.

6. An Empirical Example and Monte Carlo Experiments

This section provides some numerical results that illustrate the usefulness of our proposed tests. We use experimental data from the National Supported Work (NSW) Demonstration (LaLonde, 1986). We apply our tests to the NSW data to gain further insights into the nature of treatment effects. In addition, we carry out some Monte Carlo experiments based on the NSW data to examine the finite sample performance of our tests.

6.1. **The Data.** The NSW Demonstration (NSW) was a randomized, temporary employment program in the U.S. in the mid-1970s designed to help disadvantaged workers. A highly influential paper by LaLonde (1986) analyzed the NSW data to examine the performance of econometric evaluation estimators based on nonexperimental methods. The original sample and its subsamples were later reanalyzed by Dehejia and Wahba (1999, 2002), and Smith and Todd (2005). We use the original LaLonde (1986) sample to illustrate our proposed tests.[10] This sample consists of 297 treatment group observations and 425 control group observations. See LaLonde (1986), Dehejia and Wahba (1999, 2002), and Smith and Todd (2005) for details on the NSW data.

6.2. **Empirical Illustration.** We consider two types of outcomes Y: earnings in 1978 and changes in earnings between 1978 (postintervention year) and 1975 (preintervention year), both expressed in 1982 dollars, denoted by RE78 and RE78-RE75, respectively, as in Dehejia and Wahba (1999, 2002). The Z variable is the usual treatment indicator: $Z = 1$ for the treatment group and $Z = 0$ for the control group. There are several covariates available in the NSW data, such as age, education, earnings in 1975, and other demographic dummy variables. We use age in years as X to illustrate our tests.

Figures 1 shows nonparametric estimation results for both outcomes, RE78 and RE78-RE75. The top panel of the figure shows nonparametric estimates of conditional means of RE78 as functions of age in years (X) for the treatment and control groups, respectively. The bottom panel shows corresponding estimates for RE78-RE75. The kernel function used in estimation was

$$K(u) = \frac{3}{2}\left(1 - (2u)^2\right) 1\left\{|u| \leq \frac{1}{2}\right\}. \tag{6.1}$$

[10]The dataset is available online at Rajeev Dehejia's web page at `http://www.nber.org/ rdehejia/nswdata.html`. We thank Rajeev Dehejia for making the dataset available online.

There are several methods available for choosing a bandwidth in nonparametric kernel regression estimation. Here, we chose the bandwidth h by a simple rule of thumb, as in Section 4.2 of Fan and Gijbels (1996). To describe the rule-of-the-thumb bandwidth we used for nonparametric estimation, first note that under random assignment of treatment, $\tau_0(x) = E[Y_1 - Y_0|X = x] = E[\tilde{Y}|X = x]$, where

$$\tilde{Y} = \frac{YZ}{\Pr(Z = 1)} - \frac{Y(1 - Z)}{\Pr(Z = 0)}.$$

Now the rule-of-the-thumb bandwidth h for estimation of the average treatment effect has the following form

$$h = 3.4375 \left[\frac{\tilde{\sigma}^2 \int w_0(v)dv}{n^{-1}\sum_{i=1}^{n} \left\{ \tilde{\tau}^{(2)}(\tilde{X}_i) \right\}^2 w_0(\tilde{X}_i)} \right]^{1/5} n^{-1/5},$$

where $\tilde{X}_i$'s are studentized X_i's, $\tilde{\tau}^{(2)}(\cdot)$ is the second-order derivative of the global quartic parametric fit of $\tau_0(x)$ with studentized X_i's and with the sample proportion of Z, $\tilde{\sigma}^2$ is the simple average of squared residuals from the parametric fitting, $w_0(\cdot)$ is a uniform weight function that has value 1 for any $\tilde{X}_i$ that is between the 10th and 90th sample quantiles of $\tilde{X}$. This rule of thumb yielded the bandwidth $h = 12.679$ for RE78 and $h = 16.495$ for RE78-RE75.[11] Estimation results from Figure 1 suggest that there are positive average treatment effects for both outcomes, especially for old workers.

In Table 1, we report results from nonparametric testing. We consider four different combinations of null and alternative hypotheses:

(T1) the null of zero conditional average treatment effect (CATE) against the strictly positive CATE for some age groups (one-sided test);
(T2) the null of zero CATE against nonzero CATE for some age groups (two-sided test);
(T3) the null of the first-order stochastic dominance of the treatment group over the control group for each age group;
(T4) the null of equality between conditional distributions of treatment and control groups for all age groups.

For tests of T1 and T2, we used three weight functions described in Section 4.2: (1) the uniform weight function $w_1(x) \equiv 1$; (2) the inverse-standard-error weight function $\hat{w}_2(x) = [\hat{\rho}_2(x)]^{-1/2}$; (3) the density weight function $\hat{w}_3(x) = \hat{p}_1(x) \cdot \hat{p}_0(x)$, where all the weight functions have the support $\mathcal{W}_x$ that is an interval between the 10 and 90 percentiles of X. For tests of T3 and T4, we used similar three weight functions: (1) the uniform weight function

[11]We also applied the least cross validation to choose h and it turns out that an optimal value of h from the cross-validation was unreasonably too large.

$w_1(y,x) \equiv 1$; (2) the inverse-standard-error weight function $\hat{w}_2(y,x) = [\hat{\rho}_2(y,x)]^{-1/2}$; (3) the density weight function $\hat{w}_3(y,x) = \hat{p}_1(x) \cdot \hat{p}_0(x)$ for each y, where all the weight functions have the support $\mathcal{W}_y \times \mathcal{W}_x$. Here, $\mathcal{W}_x$ is the same as above, that is the interval between the 10 and 90 percentiles of X, and $\mathcal{W}_y$ is the entire support of Y for the uniform weight and density weight functions and the interval between the 5 and 95 percentiles of Y for the inverse-standard-error weight function, respectively.

In general, choosing a bandwidth in nonparametric testing is a difficult problem, since a good bandwidth in testing is usually different from the optimal bandwidth in estimation. If underlying functions are twice continuously differentiable, our test has a correct size and consistent for any bandwidth satisfying $C_1 n^{-C_2}$ with constants C_1 and C_2 such that $0 < C_1 < \infty$ and $1/4 < C_2 < 1/3$. To choose a bandwidth among possible values, we may need to develop a higher-order asymptotic theory based on the tradeoffs between the size and power of the test. Instead, in Table 1, we report testing results for different values of bandwidths. In particular, we considered bandwidths of the form $h = C_h \cdot \hat{s}_X \cdot n^{-2/7}$, where $\hat{s}_X$ is the sample standard deviation of the X variable, and C_h is a constant that belongs to $\{2, 2.5, 3, 3.5\}$ for the one-sided tests (T1 and T3) and $\{5, 6, 7, 8\}$ for the two-sided tests (T2 and T4). These values of the bandwidths were used in Monte Carlo experiments, which will be reported below, and seem to work reasonably well in the Monte Carlo simulations that mimic the LaLonde data.

The top panel of Table 1 displays the test result for T1. Given Figure 1, it is not surprising to find out that the null hypothesis of zero CATE is rejected in favor of a positive CATE for some age groups at the nominal level 10% across all weight functions and bandwidths. The second panel shows that the evidence is mixed if one uses the two-sided test (T2).[12] Given that we expect a positive effect from the intervention implemented in the data *a priori*, it may be more reasonable to consider T1 rather than T2. This suggests that a researcher might use a one-sided test when she expects a particular sign of the conditional average treatment effect, since the two-sided test is likely to be less powerful especially when the estimated CATE is positive for all values of X. The third panel considers the conditional stochastic dominance (T3) and shows that there is no evidence against the the null hypothesis that the control group is stochastically dominated by the treatment group for each age group. This is consistent with the result for T1. The fourth panel shows that there is no strong evidence against the equality between two conditional distributions. Again, this is in line with the result for T2.

[12]One may suspect that this may be due to the lack of the power in our nonparametric tests. However, the t-statistic for the unconditional average treatment effect ($E[Y_1 - Y_0]$) is just 1.818, which means that we fail to reject the null of zero ATE against the two-sided alternative at the 5% level.

In summary, our test results suggest that all age groups between the 10 and 90 percentiles enjoyed positive average treatment effects. This conclusion would not be made possible by just testing the statistical significance of the unconditional average treatment effect.

6.3. **Monte Carlo Experiments.** To evaluate the finite-sample performance of our tests under data generating processes (DGPs) that are similar to that of the NSW data, we treat the LaLonde (1986) sample as the true DGP in Monte Carlo experiments.

Throughout the experiments, we consider the tests for conditional average treatment effects (CATEs). In particular, we consider two tests: (i) the null hypothesis of zero CATE for every $x \in \mathcal{W}_x$ vs. the alternative hypothesis of positive CATE for some $x \in \mathcal{W}_x$ (one-sided test); (ii) the null hypothesis of zero CATE for every $x \in \mathcal{W}_x$ vs. the alternative hypothesis of nonzero CATE for some $x \in \mathcal{W}_x$ (two-sided test).

Two types of data generating processes are considered. First, 10,000 repeated samples are generated randomly with replacement from the NSW data, with the restriction that (Y, X) and Z are generated independently. We call this DGP1, which corresponds to the case that the null hypotheses in (i) and (ii) are true. Second, 10,000 repeated samples are generated randomly with replacement from the NSW data, with the joint distribution of (Y, X, Z) being left intact. We call this DGP2 and use the DGP2 to examine the powers of the tests of (i) and (ii).

In the experiments, we used three weight functions described in Section 4.2: (1) the uniform weight function $w_1(x) \equiv 1$ on $\mathcal{W}_x$; (2) the inverse-standard-error weight function $\hat{w}_2(x) = [\hat{\rho}_2(x)]^{-1/2}$; (3) the density weight function $\hat{w}_3(x) = \hat{p}_1(x) \cdot \hat{p}_0(x)$, where $\mathcal{W}_x$ is an interval between the 10 and 90 percentiles of X.

Also, we used the same kernel as in (6.1) with a bandwidth $h = C_h \hat{s}_X n^{-2/7}$, where $\hat{s}_X$ is the sample standard deviation of the X variable and C_h is a constant. In the experiments, we consider a set of different values for C_h: $\{2, 2.5, 3, 3.5\}$ for the one-sided test and $\{5, 6, 7, 8\}$ for the two-sided test. Two sample sizes were considered: $n = 722$ (the size of the original sample) and $n = 1,444$.

Tables 2 and 3 and Figure 2 summarize the results of experiments. Table 2 shows coverage probabilities of testing the null hypothesis of zero CATE for every $x \in \mathcal{W}_x$ against the alternative hypothesis of positive CATE for some $x \in \mathcal{W}_x$ (one-sided test). First of all, note that empirical rejection probabilities are not substantially different from the nominal ones with DGP1 (This is the case when H_0 is true). However, there is some tendency of overrejection for all levels of tests, especially at the 1% level tests. For DGP1, Figure 2 shows normal P-P plots for the one-sided test with $n = 722$. Each panel of the figure shows a P-P plot with a different value of the bandwidth (h). Overall, the empirical probabilities of the test statistics are similar to the normal probabilities, as asymptotic theory suggests. In addition, Table 3 and Figure 2, respectively, report coverage probabilities and the normal P-P plots of testing the null hypothesis of zero CATE for every $x \in \mathcal{W}_x$ against the alternative hypothesis of nonzero CATE for some $x \in \mathcal{W}_x$ (two-sided test). For DGP1, Monte Carlo results are similar to those for the one-sided test. For both one-sided and two-sided tests, there does not seem much difference across different weight functions.

The simulation results for DGP2 suggest that our tests are consistent since the null hypothesis is very unlikely to hold under DGP2 given our empirical analysis in Section 6.2. It seems that the power is largest with the inverse-standard-error weight function for the one-sided test and so with the uniform weight function for the two-sided test. There is no alternative test available in the literature for the one-sided test, but there exist tests for the two-sided test, for example, tests developed in Crump et al. (2008). The top panel of Table 4 shows coverage probabilities of the nonparametric test of Crump et al. (2008) with their statistic T for the null hypothesis that the conditional average treatment effect (CATE) is zero for each value of x. The bottom panel of the table shows coverage probabilities of the nonparametric test of Crump et al. (2008) with their statistic Q. It can be seen that empirical coverage probabilities of their tests are sensitive to the choice of the order of power series terms. Furthermore, it seems that our tests reported in Table 3 are more powerful or at least as powerful as their tests for most cases.

7. Conclusions

We have developed a general class of nonparametric tests for treatment effects conditional on covariates. We have shown that suitably studentized versions of our test statistics are asymptotically standard normal under the null hypotheses and have also shown that the proposed nonparametric tests are consistent against general fixed alternatives and have non-negligible powers against some local $n^{-1/2}$ alternatives.

There are several topics for further research. First, it may be an interesting research topic to develop the asymptotic properties of our tests under a more general data-generating process that goes beyond the simple random sample setup in this paper. Second, we have considered some reasonable candidates for the weight function for the test statistics. Perhaps it might be desirable to choose the weight function optimally by considering a reasonable criterion such as maximizing an average local power. Third, hypothesis testing alone might not provide a good guidance for a social planner to choose treatments (Manski, 2004). It would be an interesting topic to study whether a functional like our statistics can help the social planner to make an informed decision. Fourth, this paper does not cover marginal treatment effects that can be identified using the method of local instrumental variables developed by Heckman and Vytlacil (1999, 2005). It would be important to develop a general test for marginal treatment effects.

Appendix A. Proofs

We shall give proofs only for the test $\hat{S}$ because the proofs for the test $\hat{S}_D$ are similar and also simpler. We also omit the proof of Theorem 5.3 since it can be proved using similar arguments.

A.1. **Uniform asymptotic approximation of $\hat{T}$ by T_n.** Write

$$\hat{\tau}(y,x) = \tau_0(y,x) + [\tau_{n0}(y,x) - E\tau_{n0}(y,x)] + [E\tau_{n0}(y,x) - \tau_0(y,x)] + R_n(y,x),$$

where

$$\begin{aligned}
\tau_{n0}(y,x) &:= \frac{1}{n}\sum_{i=1}^{n} G(Y_i,y)\phi(x,Z_i)K_h(x-X_i)\,, \\
R_n(y,x) &:= \frac{1}{n}\sum_{i=1}^{n} G(Y_i,y)\phi(x,Z_i) \\
&\quad \times \left[1(Z_i=1)\frac{p_1(x)-\hat{p}_1(x)}{\hat{p}_1(x)} + 1(Z_i=0)\frac{p_0(x)-\hat{p}_0(x)}{\hat{p}_0(x)}\right] K_h(x-X_i)\,.
\end{aligned}$$

Define

$$\begin{aligned}
\zeta_n(y,x) &= E[G(Y,y)|X=x,Z=1] - E[G(Y,y)|X=x,Z=0] \\
&\quad - E[G(Y,y)|X=x,Z=1]\frac{1}{np_1(x)}\sum_{i=1}^{n} 1(Z_i=1)K_h(x-X_i) \\
&\quad + E[G(Y,y)|X=x,Z=0]\frac{1}{np_0(x)}\sum_{i=1}^{n} 1(Z_i=0)K_h(x-X_i)\,.
\end{aligned}$$

The following lemma shows that $R_n(y,x)$ can be approximated by $\zeta_n(y,x)$ uniformly over (y,x) at a rate faster than $n^{-1/2}$.

Lemma A.1. *Under Assumption 4.1, we have that*

$$\sup_{(y,x)\in\mathcal{W}} |R_n(y,x) - \zeta_n(y,x)| = o_p(n^{-1/2}).$$

Proof of Lemma A.1. Note that under the conditions on the bandwidth,

$$\max_{x\in\mathcal{W}_x} |\hat{p}_j(x) - p_j(x)| = O_p\left[h^s + (nh^d)^{-1/2}(\log n)^{1/2}\right] = o_p(n^{-1/4})$$

for $j=0,1$. Then the following holds uniformly over (y,x):

$$R_n(y,x) = R_{n1}(y,x) + R_{n2}(y,x) + R_{n3}(y,x) + o_p\left(n^{-1/2}\right),$$

where

$$\begin{aligned}
R_{n1}(y,x) &:= \frac{1}{n}\sum_{i=1}^{n} G(Y_i,y)\phi(x,Z_i)K_h(x-X_i)\,, \\
R_{n2}(y,x) &:= \frac{1}{n^2 p_1^2(x)}\sum_{i=1}^{n}\sum_{j=1}^{n} \zeta_1(W_i,W_j,y,x),
\end{aligned}$$

$$R_{n3}(y,x) := \frac{1}{n^2 p_0^2(x)} \sum_{i=1}^{n} \sum_{j=1}^{n} \zeta_0(W_i, W_j, y, x),$$

with

$$\zeta_1(W_i, W_j, y, x) := -G(Y_i, y) 1(Z_i = 1) 1(Z_j = 1) K_h(x - X_i) K_h(x - X_j),$$
$$\zeta_0(W_i, W_j, y, x) := G(Y_i, y) 1(Z_i = 0) 1(Z_j = 0) K_h(x - X_i) K_h(x - X_j).$$

Split $R_{n2}(y,x)$ as

$$R_{n2}(y,x) = \frac{1}{n^2 p_1^2(x)} \sum_{i=1}^{n} \sum_{j=1, j\neq i}^{n} \zeta_1(W_i, W_j, y, x) + \frac{1}{n^2 p_1^2(x)} \sum_{i=1}^{n} \zeta_1(W_i, W_i, y, x), \tag{A.1}$$

where the second term has the form

$$\frac{1}{n^2 p_1^2(x)} \sum_{i=1}^{n} \zeta_1(W_i, W_i, y, x) = -\frac{1}{n^2 h^{2d} p_1^2(x)} \sum_{i=1}^{n} G(Y_i, y) 1(Z_i = 1) K^2\left(\frac{x - X_i}{h}\right).$$

Since K is of bounded variation and $\{G(\cdot, y) : y \in \mathcal{W}_y\}$ is a VC class, standard results in empirical process methods (see. e.g. Theorem 2.14.1 of van der Vaart and Wellner (1996, p.239)) yield

$$\sup_{(y,x)\in\mathcal{W}} \left| \frac{1}{nh^d} \sum_{i=1}^{n} G(Y_i, y) 1(Z_i = 1) K^2\left(\frac{x - X_i}{h}\right) \right| = O_p(1),$$

which implies that

$$\sup_{(y,x)\in\mathcal{W}} \left| \frac{1}{n^2 p_1^2(x)} \sum_{i=1}^{n} \zeta(W_i, W_i, y, x) \right| = O_p\left[(nh^d)^{-1}\right]$$

since $p_1(\cdot)$ is bounded away from zero on $\mathcal{W}_x$.

We now move on the first term in (A.1). We will apply the uniform approximation result for U-processes (see, e.g. Ghosal et al., 2000). To do so, let $\mathbb{U}_n$ denote the random discrete measure putting mass $1/n(n-1)$ for each of the points $\{(W_i, W_j) : 1 \leq i < j \leq n\}$. Note that

$$\begin{aligned}\frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1, j\neq i}^{n} \zeta_1(W_i, W_j, y, x) &= \frac{1}{n^2} \sum_{1\leq i<j\leq n} \zeta_1(W_i, W_j, y, x) + \zeta_1(W_j, W_i, y, x) \\ &= \mathbb{U}_n \tilde{\zeta}_{(y,x)} [1 + o_p(1)],\end{aligned}$$

where $\tilde{\zeta}_{(y,x)}(W_i, W_j) = \zeta_1(W_i, W_j, y, x) + \zeta_1(W_j, W_i, y, x)$.

Consider a class of functions

$$\mathcal{F} = \{\tilde{\zeta}_{(y,x)} : (y,x) \in \mathcal{W}\}.$$

Note that $\mathcal{F}$ is contained in $\mathcal{F}_1 \times \mathcal{F}_2 \times \mathcal{F}_3 + \mathcal{F}_4 \times \mathcal{F}_2 \times \mathcal{F}_3$, where

$$\begin{aligned}\mathcal{F}_1 &= \{G(Y_i, y) : y \in \mathcal{W}_y\}, \\ \mathcal{F}_2 &= \left\{K\left(\frac{x - X_i}{h}\right) K\left(\frac{x - X_j}{h}\right) : x \in \mathcal{W}_x\right\}, \\ \mathcal{F}_3 &= h^{-2d} 1(Z_i = 1) 1(Z_j = 1) 1(\|X_i - X_j\| \leq h),\end{aligned}$$

$$\mathcal{F}_4 = \{G(Y_j, y) : y \in \mathcal{W}_y\}.$$

By Assumption, $\mathcal{F}_1$ and $\mathcal{F}_4$ are VC classes of functions with the envelope function $\mathbf{M}$. Since K is of bounded variation, $\mathcal{F}_2$ is also a VC class of functions with the bounded envelope function. Note that $\mathcal{F}_3$ is not indexed by (y, x) and is bounded by $h^{-2d}1(\|X_i - X_j\| \leq h)$. Hence, we can take an envelope function $\mathbf{F}$ of $\mathcal{F}$ to be $[\mathbf{M}(Y_i) + \mathbf{M}(Y_j)]h^{-2d}1(\|X_i - X_j\| \leq h)$.

Let

$$\widehat{U}_n\tilde{\zeta}_{(y,x)} = n^{-1}\sum_{i=1}^{n} E[\zeta_1(W_i, W_j, y, x)|W_i] + n^{-1}\sum_{j=1}^{n} E[\zeta_1(W_j, W_i, y, x)|W_j] - E[\zeta_1(W_j, W_i, y, x)].$$

Then by Theorem A.1 of Ghosal et al. (2000) and comments following this theorem, there exists a universal constant $C < \infty$ and such that

$$\begin{aligned} &E\left(\sup\left\{\mathbb{U}_n\tilde{\zeta}_{(y,x)} - \widehat{U}_n\tilde{\zeta}_{(y,x)}| : \tilde{\zeta}_{(y,x)} \in \mathcal{F}\right\}\right) \\ &\leq Cn^{-1}(E\mathbf{F}^2)^{1/2}\int_0^1 \sup_Q \log N\left(\varepsilon\left\|\mathbf{F}\right\|_{Q,2}, \mathcal{F}, L_2(Q)\right) d\varepsilon, \end{aligned}$$

where $N(\varepsilon, \mathcal{F}, L_2(Q))$ is the ε-covering number of $\mathcal{F}$ with the $L_2(Q)$ norm. Here, Q denotes a probability. Note that

$$(E\mathbf{F}^2)^{1/2} \leq Ch^{-3d/2}.$$

Furthermore, by Lemma A.1 of Ghosal et al. (2000) and also by the fact that the 2ε-covering numbers of the sum of the two classes are bounded by the product of the ε-covering numbers of the two classes,

$$\int_0^1 \sup_Q \log N\left(\varepsilon\left\|\mathbf{F}\right\|_{Q,2}, \mathcal{F}, L_2(Q)\right) d\varepsilon \leq C\int_0^1 \log\varepsilon^{-1} d\epsilon < \infty.$$

Then, combining results above with the bandwidth requirement that $nh^{3d} \to \infty$ gives

$$\sup_{(y,x)\in\mathcal{W}}\left|\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1,j\neq i}^{n}\zeta_1(W_i, W_j, y, x) - \widehat{U}_n\tilde{\zeta}_{(y,x)}\right| = o_p(n^{-1/2}). \tag{A.2}$$

By standard arguments for kernel estimation,

$$\begin{aligned} E[\zeta_1(W_i, W_j, y, x)|W_i] &= -G(Y_i, y)1(Z_i = 1)1(Z_j = 1)K_h(x - X_i)\,p_1(x) + O(h^s), \\ E[\zeta_1(W_i, W_j, y, x)|W_j] &= -E[G(Y, y)|X = x, Z = 1]p_1(x)1(Z_j = 1)K_h(x - X_j) + O(h^s), \\ E[\zeta_1(W_i, W_j, y, x)] &= -E[G(Y, y)|X = x, Z = 1]p_1^2(x) + O(h^s) \end{aligned}$$

uniformly over (y, x). Thus, combining the results above with (A.1)-(A.2) gives

$$\begin{aligned} R_{n2}(y, x) = &-\frac{1}{n}\sum_{i=1}^{n}\{G(Y_i, y) + E[G(Y, y)|X = x, Z = 1]\}\frac{1(Z_i = 1)}{p_1(x)}K_h(x - X_i) \\ &+ E[G(Y, y)|X = x, Z = 1] + o_p(n^{-1/2}) \end{aligned}$$

uniformly over (y, x). Arguments identical to prove the equation above yield

$$\begin{aligned} R_{n3}(y,x) = &\frac{1}{n}\sum_{i=1}^{n}\{G(Y_i,y) + E[G(Y,y)|X=x, Z=0]\}\frac{1(Z_i=0)}{p_0(x)}K_h(x-X_i) \\ &- E[G(Y,y)|X=x, Z=0] + o_p(n^{-1/2}) \end{aligned}$$

uniformly over (y, x). Therefore, combining results all together proves the lemma. ∎

Now define

$$T_n^* := \int\int \sqrt{n}\max\{\tau_0(y,x) + [\tau_n(y,x) - E\tau_n(y,x)], 0\}w(y,x)dydx, \tag{A.3}$$

where

$$\tau_n(y,x) := \tau_{n0}(y,x) + \zeta_n(y,x). \tag{A.4}$$

Lemma A.2. *Under Assumption 4.1, we have that*

$$\hat{T} = T_n^* + o_p(1).$$

Proof of Lemma A.2. Since $|\max\{a,0\} - \max\{b,0\}| \leq |a-b|$, we have

$$\begin{aligned} &|\hat{T} - T_n^*| \\ &\leq \int\int \sqrt{n}|E\tau_{n0}(y,x) - \tau_0(y,x)|w(y,x)dydx \\ &+ \int\int \sqrt{n}|E\zeta_n(y,x)|w(y,x)dydx + \int\int \sqrt{n}|R_n(y,x) - \zeta_n(y,x)|w(y,x)dydx. \end{aligned}$$

By Lemma A.1, the third term above is asymptotically negligible. Also, by Taylor's Theorem and standard arguments for kernel estimation along with the fact that K is a s-order kernel and $nh^{2s} \to 0$ as $n \to \infty$, we have that

$$\begin{aligned} &\int\int \sqrt{n}|E\zeta_n(y,x)|w(y,x)dydx = O\left(n^{1/2}h^s\right) = o(1), \\ &\int\int \sqrt{n}|E\tau_{n0}(y,x) - \tau_0(y,x)|w(y,x)dydx = O\left(n^{1/2}h^s\right) = o(1). \end{aligned}$$

Thus, we have proved the lemma. ∎

Hence, under the null hypothesis that $\tau_0(y,x) \equiv 0$ on $\mathcal{W}$, we have that $\hat{T} = T_n + o_p(1)$, where T_n was defined in (4.10).

A.2. **The Asymptotic Null Distribution.** We first establish that the estimators of the asymptotic bias and variance are consistent. To do so, the following lemmas are useful.

Lemma A.3. *Under Assumption 4.1, the following holds:*

$$\text{(a)} \sup_{(y,x)\in\mathcal{W}} |\hat{\tau}(y,x)-\tau_0(y,x)| = O_p\left[\left(nh^d\right)^{-1/2}\log n + h^s\right],$$

$$\text{(b)} \sup_{(y,x)\in\mathcal{W}} |\hat{\rho}_2(y,x)-\rho_2(y,x)| = O_p\left[\left(nh^d\right)^{-1/2}\log n + h^s\right].$$

Proof of Lemma A.3. We first verify (a). Write

$$\begin{aligned}\hat{\tau}(y,x) &= \tau_0(y,x) + [\tau_{n0}(y,x) - E\tau_{n0}(y,x)] + [E\tau_{n0}(y,x) - \tau_0(y,x)] \\ &\quad + [\zeta_n(y,x) - E\zeta_n(y,x)] + E\zeta_n(y,x) \\ &\quad + [R_n(y,x) - \zeta_n(y,x)].\end{aligned}$$

Repeated applications of Theorem 37 of Pollard (1984, p.34) give

$$\sup_{(y,x)\in\mathcal{W}} |\tau_{n0}(y,x) - E\tau_{n0}(y,x)| = O\left[\left(nh^d\right)^{-1/2}\log n\right], \tag{A.5}$$

$$\sup_{(y,x)\in\mathcal{W}} |\zeta_n(y,x) - E\zeta_n(y,x)| = O\left[\left(nh^d\right)^{-1/2}\log n\right] \tag{A.6}$$

almost surely. Also, note that by usual bias calculations in kernel estimation,

$$E\tau_{n0}(y,x) - \tau_0(y,x) = O(h^s) \text{ and } E\zeta_{n0}(y,x) = O(h^s)$$

uniformly over (y,x). Then part (a) follows from Lemma A.1. The proof of part (b) is similar. ∎

Theorem A.1. *Under Assumption 4.1, we have*

$$\begin{aligned}\text{(a)}\ \hat{a}_n &= a_n + o_p(1),\\ \text{(b)}\ \hat{\sigma}^2 &= \sigma_0^2 + o_p(1).\end{aligned}$$

Proof of Theorem A.1. Note that we have

$$\begin{aligned}&\left|\int \sqrt{\hat{\rho}_2(y,x)}w(y,x)du - \int \sqrt{\rho_2(y,x)}w(y,x)du\right| \\ \le\ & \int \left|\sqrt{\hat{\rho}_2(y,x)} - \sqrt{\rho_2(y,x)}\right| w(y,x)du \\ \le\ & \left[\inf_{(y,x)\in\mathcal{W}} |\rho_2(y,x)|\right]^{-1/2} \sup_{(y,x)\in\mathcal{W}} |\hat{\rho}_2(y,x)-\rho_2(y,x)| \\ =\ & O_p(n^{-1/2}h^{-d/2}\log n + h^s) = o_p(h^{d/2})\end{aligned}$$

where the first inequality holds by triangle inequality, the second inequality holds by the simple inequality $\left|\sqrt{a}-\sqrt{b}\right| = \left|\sqrt{a}+\sqrt{b}\right|^{-1}|a-b| \le a^{-1/2}|a-b|$ for $a,b>0$, the first equality holds by Lemma A.3 and Assumption 4.1 (iv) and (vi), and the last equality holds by Assumption 4.1 (ix). This establishes part (a) of Theorem A.1. The proof of part (b) is similar since

$$F(\rho) = Cov\left(\max\{\sqrt{1-\rho}\mathbb{Z}_1 + \rho\mathbb{Z}_2, 0\}, \max\{\mathbb{Z}_2, 0\}\right)$$

is a continuous functional of ρ on T_0 and $\hat{\rho}^2(\cdot,\cdot,\cdot,\cdot)$ is consistent for $\rho^2(\cdot,\cdot,\cdot,\cdot)$ uniformly over $\mathcal{W}_y \times \mathcal{W}_y \times \mathcal{W}_x \times T_0$ using Lemma A.3 and Assumption 4.1 (iv) and (vi). ∎

We need to show that the asymptotic distribution of T_n is normal:

Theorem A.2. *Under Assumption 4.1, we have*

$$\frac{T_n - a_n}{\sigma_0} \xrightarrow{d} N(0,1).$$

The proof of Theorem A.2 is lengthy and will be given below in Section A.3. Given Theorems A.1 and A.2, we can establish Theorem 4.1.

Proof of Theorem 4.1. We have

$$\begin{aligned} \Pr\left(\hat{S} > z_{1-\alpha}\right) &= \Pr\left(\hat{T} > \hat{a}_n + \hat{\sigma} z_{1-\alpha}\right) \\ &= \Pr\left(T_n^* > \hat{a}_n + \hat{\sigma} z_{1-\alpha}\right) + o(1) \\ &\leq \Pr\left(T_n > \hat{a}_n + \hat{\sigma} z_{1-\alpha}\right) + o(1) \\ &\to \alpha, \end{aligned}$$

where the second equality holds by Lemma A.2, the inequality holds since $\tau_0(y,x) \leq 0$ for each $(y,x) \in \mathcal{W}$ under the null hypothesis (with inequality replaced by equality if $\tau_0(y,x) = 0$ for each $(y,x) \in \mathcal{W}$), and the last convergence to α follows from Theorems A.1 and A.2. This gives the desired result of Theorem 4.1. ∎

A.3. **Proof of Theorem A.2.** We now establish Theorem A.2. For this purpose, we need several lemmas. The first lemma is related to the Berry-Esseen theorem.

Lemma A.4. *Let $\{\tilde{W}_i = (\tilde{W}_{1i}, \tilde{W}_{2i})' : i \geq 1\}$ be a sequence of i.i.d. random vectors in $\mathbb{R}^2$ such that each component has mean 0, variance 1, and finite absolute moments of third order. Let $\bar{Z} = (\bar{Z}_1, \bar{Z}_2)'$ be multivariate normal with mean vector 0 and variance-covariance matrix*

$$\Sigma = E\bar{Z}\bar{Z}' = E\tilde{W}\tilde{W}' = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}.$$

Let μ_1 and μ_2 be finite constants. Then there exist universal positive constants A_1, A_2, A_2', A_3 and A_3' such that

$$\left| E\max\left\{\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\tilde{W}_{1i} + \mu_1, 0\right\} - E\max\left\{\bar{Z}_1 + \mu_1, 0\right\}\right| \leq \frac{A_1}{\sqrt{n}} E\left|\tilde{W}_1\right|^3 \tag{A.7}$$

and, whenever $\rho^2 < 1$,

$$\begin{aligned}
&\left| E\left[\max\left\{\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\tilde{W}_{1i}+\mu_1,0\right\}\max\left\{\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\tilde{W}_{2i}+\mu_2,0\right\}\right]\right.\\
&\qquad\left. - E\left[\max\left\{\bar{Z}_1+\mu_1,0\right\}\max\left\{\bar{Z}_2+\mu_2,0\right\}\right]\right| \qquad \text{(A.8)}\\
&\le \frac{A_2}{(1-\rho^2)^{3/2}}\frac{\log n}{\sqrt{n}}\left(E\left|\tilde{W}_1\right|^3+E\left|\tilde{W}_2\right|^3\right)+\frac{A_2'}{(1-\rho^2)^3}\frac{(\log n)^2}{n}\left(E\left|\tilde{W}_1\right|^3+E\left|\tilde{W}_2\right|^3\right)^2
\end{aligned}$$

and

$$\begin{aligned}
&\left| E\left[\max\left\{\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\tilde{W}_{1i}+\mu_1,0\right\}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\tilde{W}_{2i}\right] - E\left[\max\left\{\bar{Z}_1+\mu_1,0\right\}\bar{Z}_2\right]\right| \qquad \text{(A.9)}\\
&\le \frac{1}{(1-\rho^2)^{3/2}}\frac{A_3}{\sqrt{n}}\left(E\left|\tilde{W}_1\right|^3+E\left|\tilde{W}_2\right|^3\right)+\frac{A_3'}{(1-\rho^2)^3}\frac{(\log n)^2}{n}\left(E\left|\tilde{W}_1\right|^3+E\left|\tilde{W}_2\right|^3\right)^2.
\end{aligned}$$

Proof of Lemma A.4. We prove this lemma using the results in Bhattacharya (1975). In particular, special cases of the main theorem of Bhattacharya (1975) provide the following facts.

Fact A.1. *Let $\|\cdot\|$ denote the Euclidean norm in $\mathbb{R}^k$. Let $\tilde{X}_1,\ldots,\tilde{X}_n$ be n independent and identically distributed random vectors in $\mathbb{R}^k$ with $E\tilde{X}_1=0$ and $Cov(\tilde{X}_1)=I$, where I is the identity matrix. Let $\mathbb{Z}$ be a vector of independent standard normals in $\mathbb{R}^k$.*

(a) Assume that $k=1$ and that a function $h:\mathbb{R}\mapsto\mathbb{R}$ satisfies

$$|h(x)-h(y)|\le C_1\|x-y\|,\quad \sup_{x\in\mathbb{R}}\frac{|h(x)|}{1+\|x\|^r}\le C_2$$

for some positive, finite constants C_1 and C_2 and some integer r, $0\le r\le 3$. Then there exists a universal constant $C_3<\infty$ such that

$$\left|E\left[h\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\tilde{X}_i\right)\right]-E\left[h(\mathbb{Z})\right]\right|\le C_3E\left\|\tilde{X}_1\right\|^3n^{-1/2}.$$

(b) Now assume that $k=2$ and that a function $h:\mathbb{R}^2\mapsto\mathbb{R}$ satisfies

(i) *$h(x_1,x_2)=h_1(x_1)h_2(x_2)$;*
(ii) *$|h_j(x_j)-h_j(y_j)|\le C_1|x_j-y_j|$ for $j=0,1$;*
(iii) *The following holds uniformly in (y_1,y_2) such that $\|(x_1,x_2)-(y_1,y_2)\|\le\varepsilon$:*

$$\sup_{(y_1,y_2):\|(x_1,x_2)-(y_1,y_2)\|\le\varepsilon}|h(x_1,x_2)-h(y_1,y_2)|\le C_1\left\{\varepsilon\left[|h_1(x_1)|+|h_2(x_2)|\right]+\varepsilon^2\right\};$$

(iv) *$\sup_{(x_1,x_2)\in\mathbb{R}^2}(1+\|(x_1,x_2)\|^r)^{-1}|h(x_1,x_2)|\le C_2$*

for some positive, finite constants C_1 and C_2 and some integer r, $0\le r\le 3$. Then there exists a universal constant $C_3<\infty$ such that

$$\left|E\left[h\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\tilde{X}_i\right)\right]-E\left[h(\mathbb{Z})\right]\right|$$

$$\leq C_3\left[E\left\|\tilde{X}_1\right\|^3 n^{-1/2} + E\left\|\tilde{X}_1\right\|^3 n^{-1/2}\log n\, E\left\{|h_1(\mathbb{Z}^{(1)})| + |h_2(\mathbb{Z}^{(2)})|\right\} + \left\{E\left\|\tilde{X}_1\right\|^3\right\}^2 n^{-1}(\log n)^2\right],$$

where $\mathbb{Z}^{(j)}$ *is the* j*-th element of* $\mathbb{Z}$.

Fact A.1 (a) comes from Section 2.2 of Bhattacharya (1975) and Fact A.1 (b) follows from equation (1.11) of Bhattacharya (1975). Now the first conclusion (A.7) of the lemma follows directly from Fact A.1 (a) with $h(x) = \max\{x + \mu_1, 0\}$ and $r = 1$ since

$$|\max\{x + \mu_1, 0\} - \max\{y + \mu_1, 0\}| \leq |x - y| \text{ and } \frac{|\max\{x + \mu_1, 0\}|}{1 + |x|} \leq 1 + |\mu_1|.$$

To show the second conclusion (A.8) of the lemma, let $h(x_1, x_2) = \max\{x_1 + \mu_1, 0\}\max\{x_2 + \mu_2, 0\}$. Then it is straightforward to show conditions (i)-(iii). For condition (iv), choose $r = 2$. Note that

$$\begin{aligned}
\frac{|h(x_1, x_2)|}{1 + |x_1|^2 + |x_2|^2} &\leq \frac{|x_1 + \mu_1||x_2 + \mu_2|}{1 + |x_1|^2 + |x_2|^2} \\
&\leq \frac{|x_1||x_2|}{1 + (|x_1| - |x_2|)^2 + 2|x_1||x_2|} + \frac{|\mu_1||\mu_2|}{1 + |x_1|^2 + |x_2|^2} \\
&\quad + |\mu_1|\frac{|x_1|^2 1\{|x_1| \geq 1\} + 1\{|x_1| < 1\}}{1 + |x_1|^2 + |x_2|^2} + |\mu_2|\frac{|x_2|^2 1\{|x_2| \geq 1\} + 1\{|x_2| < 1\}}{1 + |x_1|^2 + |x_2|^2} \\
&\leq 1 + |\mu_1||\mu_2| + 2|\mu_1| + 2|\mu_2|.
\end{aligned}$$

Hence, we have verified condition (iv). Then as long as $\rho^2 < 1$, (A.8) follows from Fact A.1 (b) using the change of variables based on $\tilde{X} = \Sigma^{-1/2}\tilde{W}$. The third conclusion (A.9) of the lemma can be proved using arguments similar to those used in the proof of (A.8). ∎

We omit the proof of the following Lemma since it is similar to that of Lemma 6.1 of Giné et al. (2003).

Lemma A.5. *Suppose* $\mathcal{H}$ *is a finite class of uniformly bounded functions* $H : \mathbb{R}^d \to \mathbb{R}$, *which are equal to zero outside of a compact set. Also, suppose* $g(y, x)f(x)$ *is continuously differentiable in* x *with* $\sup_{(y,x)\in B} |D_x g(y, x)f(x)| < \infty$, *where* $B \subset \mathbb{R}^{d+1}$ *is a compact set. Then, uniformly in* $H \in \mathcal{H}$,

$$\sup_{(y,x)\in B}\left|\frac{1}{h^d}\int_{-\infty}^{\infty} g(y, z)f(z)H\left(\frac{x - z}{h}\right)dz - J(H)g(y, x)f(x)\right| \to 0 \text{ as } h \to 0, \tag{A.10}$$

where

$$J(H) = \int_{\mathbb{R}^d} H(u)du.$$

We prove Theorem A.2 by extending the "Poissonization" result of Giné et al. (2003). We first introduce some concepts used throughout the proof. Let N denote a Poisson random variable with mean n, defined on the same probability space as the sequence $\{W_i : i \geq 1\}$ and independent of this sequence. Define

$$\chi_j := \frac{E[G(Y,y)|X=x,Z=j]}{p_j(x)},$$
$$\chi(z,y,x) := \chi_1(y,x)1(z=1) - \chi_0(y,x)1(z=0),$$
$$\varphi(W_i,y,x) := [G(Y_i,y)\phi(x,Z_i) - \chi(Z_i,y,x)]\, K_h(x-X_i) + \tau_0(y,x).$$

Recall that $\tau_n(y,x)$ is defined in (A.4). Then

$$\tau_n(y,x) = \tau_{n0}(y,x) + \zeta_n(y,x) = \frac{1}{n}\sum_{i=1}^{n}\varphi(W_i,y,x).$$

Now we will Poissonize $\tau_n(y,x)$. To do so, define

$$\tau_N(y,x) = \frac{1}{n}\sum_{i=1}^{N}\varphi(W_i,y,x), \tag{A.11}$$

where the empty sum is defined to be zero. Notice that

$$E\tau_N(y,x) = E\tau_n(y,x) = E\left[\varphi(W,y,x)\right], \tag{A.12}$$
$$k_{\tau,n}(y,x) := nVar\left(\tau_N(y,x)\right) = E\left[\varphi^2(W,y,x)\right], \tag{A.13}$$
$$nVar\left(\tau_n(y,x)\right) = E\left[\varphi^2(W,y,x)\right] - \left\{E\left[\varphi(W,y,x)\right]\right\}^2. \tag{A.14}$$

Let $\varepsilon \in \left(0, \int_{\mathcal{W}_x} f(x)dx\right)$ be an arbitrary constant. For constant $\{M_j > 0 : j = 1,...,d\}$, let $\mathcal{B}(M) = \prod_{j=1}^{d}[-M_j, M_j] \subset \mathcal{W}_x$ denote a Borel set in $\mathbb{R}^d$ with nonempty interior with finite Lebesgue measure $\lambda(\mathcal{B}(M))$. For $v > 0$, define $\mathcal{B}(M,v)$ to be the v-contraction of $\mathcal{B}(M)$, i.e., $\mathcal{B}(M,v) = \{x \in \mathcal{B}(M) : \rho(x, \mathbb{R}^d \backslash \mathcal{B}(M)) \geq v\}$, where $\rho(x,B) = \inf\{\|x-y\| : y \in B\}$. Choose $M, v > 0$ and a Borel set B_0 such that

$$B_0 \subset \mathcal{B}(M,v), \tag{A.15}$$
$$\int_{\mathbb{R}^d \backslash \mathcal{B}(M)} f(x)dx := \alpha > 0, \tag{A.16}$$
$$\int_{B_0} f(x)dx > \int_{\mathcal{W}_x} f(x)dx - \varepsilon. \tag{A.17}$$

Such M, ν, and B_0 exist by the absolute continuity of the density f, see also Lemma 6.1 of Giné et al. (2003).

Let $B = \mathbb{R} \times B_0$ and define a Poissonization version of T_n (restricted to B) to be:

$$\begin{aligned} T_n^P(B) \;=\; & \int_{B_0}\int_{\mathbb{R}} \sqrt{n}\max\{\left[\tau_N(y,x) - E\tau_n(y,x)\right], 0\}w(y,x)dydx \\ & - \int_{B_0}\int_{\mathbb{R}} \sqrt{n}E\max\{\left[\tau_N(y,x) - E\tau_n(y,x)\right], 0\}w(y,x)dydx. \end{aligned} \tag{A.18}$$

Also, let

$$\sigma_n^2(B) = Var\left(T_n^P(B)\right). \tag{A.19}$$

The following lemma derives the asymptotic variance of $T_n^P(B)$.

Lemma A.6. *If Assumption 4.1 holds and B satisfies (A.15)-(A.17), we have*

$$\lim_{n\to\infty} \sigma_n^2(B) = \sigma_{0,B}^2, \tag{A.20}$$

where

$$\begin{aligned} \sigma_{0,B}^2 \;=\; & \int_{T_0}\int_{B_0}\int_{\mathbb{R}}\int_{\mathbb{R}} Cov\left(\max\{\sqrt{1-\rho^2(y,y',x,t)}Z_1 + \rho(y,y',x,t)Z_2, 0\}, \max\{Z_2,0\}\right) \\ & \times \sqrt{\rho_2(y,x)\rho_2(y',x)}w(y,x)w(y',x)dydy'dxdt. \end{aligned} \tag{A.21}$$

Proof of Lemma A.6. To show (A.20), notice that, for each $(y,x),(y',x') \in \mathbb{R}^{d+1}$ such that $\|x-x'\| > h$, the random variables $\tau_N(y,x) - E\tau_n(y,x)$ and $\tau_N(y',x') - E\tau_n(y',x')$ are independent because they are functions of independent increments of a Poisson process and the kernel K vanishes outside of the closed ball of radius $1/2$. Therefore,

$$\begin{aligned} & Var\left(T_n^P(B)\right) \\ & = n\int_{B_0}\int_{B_0}\int_{\mathbb{R}}\int_{\mathbb{R}} Cov\left(\max\{[\tau_N(y,x) - E\tau_n(y,x)], 0\}, \max\{\left[\tau_N(y',x') - E\tau_n(y',x')\right], 0\}\right) \\ & \times w(y,x)w(y',x')dydy'dxdx' \\ & = n\int_{B_0}\int_{B_0}\int_{\mathbb{R}}\int_{\mathbb{R}} 1\left(\|x-x'\| \le h\right) \\ & \times Cov\left(\max\{[\tau_N(y,x) - E\tau_n(y,x)], 0\}, \max\{\left[\tau_N(y',x') - E\tau_n(y',x')\right], 0\}\right) \\ & \times w(y,x)w(y',x')dydy'dxdx'. \end{aligned}$$

Let

$$S_{\tau,N}(y,x) = \frac{\sqrt{n}\left\{\tau_N(y,x) - E\tau_n(y,x)\right\}}{\sqrt{k_{\tau,n}(y,x)}}, \tag{A.22}$$

where $k_{\tau,n}(y,x) = nVar\left(\tau_N(y,x)\right)$ is given by (A.13). We have that, with $\lambda(\mathcal{W}_y \times B_0) < \infty$,

$$\sup_{(y,x)\in\mathcal{W}_y\times B_0} \left|\sqrt{k_{\tau,n}(y,x)} - h^{-d/2}\sqrt{\rho_2(y,x)}\right| = O\left(h^{d/2}\right), \tag{A.23}$$

$$\int_{B_0}\int_{B_0}\int_{\mathbb{R}}\int_{\mathbb{R}} 1\left(\|x-x'\| \le h\right) w(y,x)w(y',x')dydy'dxdx' = O(h^d), \tag{A.24}$$

$$\sup_{\{(y,x),(y',x')\}\in(\mathcal{W}_y\times B_0)^2} \left|Cov\left(\max\{S_{\tau,N}(y,x),0\}, \max\{S_{\tau,N}(y',x'),0\}\right)\right| = O(1), \tag{A.25}$$

where (A.23) holds by Lemma A.5 and (A.25) follows from Cauchy Schwartz inequality. Therefore, from (A.23) - (A.25), we have that

$$Var\left(T_n^P(B)\right) = \overline{\sigma}_{n,0}^2 + o(1),$$

where

$$\overline{\sigma}_{n,0}^2 = \int_{B_0}\int_{B_0}\int_{\mathbb{R}}\int_{\mathbb{R}} 1\left(\|x-x'\| \le h\right) Cov\left(\max\{S_{\tau,N}(y,x),0\},\max\{S_{\tau,N}(y',x'),0\}\right)$$
$$\times h^{-d}\sqrt{\rho_2(y,x)\rho_2(y',x')}w(y,x)w(y',x')dydy'dxdx'. \tag{A.26}$$

Now, let $(Z_{1n}(y,x), Z_{2n}(y',x'))$ for $(y,x),(y',x') \in \mathbb{R}^{d+1}$, be a mean zero multivariate Gaussian process such that, for each $(y,x) \in \mathbb{R}^{d+1}$ and $(y',x') \in \mathbb{R}^{d+1}$, $(Z_{1n}(y,x), Z_{2n}(y',x'))$ and $(S_{\tau,N}(y,x), S_{\tau,N}(y',x'))$ have the same covariance structure. That is,

$$\begin{aligned}&\left(Z_{1n}(y,x), Z_{2n}(y',x')\right)\\ &\overset{d}{=} \left(\sqrt{1-(\rho_n^*(y,y',x,x'))^2}\mathbb{Z}_1 + \rho_n^*(y,y',x,x')\mathbb{Z}_2, \mathbb{Z}_2\right),\end{aligned}$$

where $\mathbb{Z}_1$ and $\mathbb{Z}_2$ are independent standard normal random variables and

$$\rho_n^*(y,y',x,x') = E\left[S_{\tau,N}(y,x)S_{\tau,N}(y',x')\right].$$

Let

$$\overline{\tau}_{n,0}^2 = \int_{B_0}\int_{B_0}\int_{\mathbb{R}}\int_{\mathbb{R}} 1\left(\|x-x'\| \le h\right) Cov\left(\max\{Z_{1n}(y,x),0\},\max\left\{Z_{2n}(y',x'),0\right\}\right)$$
$$\times h^{-d}\sqrt{\rho_2(y,x)\rho_2(y',x')}w(y,x)w(y',x')dydy'dxdx'.$$

By a change of variables $x' = x+th$, we can write

$$\overline{\tau}_{n,0}^2 = \int_{T_0}\int_{B_0}\int_{\mathbb{R}}\int_{\mathbb{R}} 1(x\in B_0)1\left(x+th \in B_0\right) Cov\left(\max\{Z_{1n}(y,x),0\},\max\left\{Z_{2n}(y',x+th),0\right\}\right)$$
$$\times\sqrt{\rho_2(y,x)\rho_2(y',x+th)}w(y,x)w(y',x+th)dydy'dxdt.$$

Note that

$$\begin{aligned}&nE\left[\{\tau_N(y,x)-E\left[\tau_N(y,x)\right]\}\left\{\tau_N(y',x')-E\left[\tau_N(y',x')\right]\right\}\right]\\ &= E\left\{\left[G(Y,y)\phi(x,Z)-\chi(Z,y,x)\right]\left[G(Y,y')\phi(x',Z)-\chi(Z,y',x')\right]K_h\left(x-X\right)K_h\left(x'-X\right)\right\}.\end{aligned}$$

Then, by Lemma A.5 and a change of variables $x' = x+th$, we have, for almost every (y,y',x,t),

$$\rho_n^*(y,y',x,x+th) \to \frac{\rho_1(y,y',x,t)}{\sqrt{\rho_2(y,x)\rho_2(y',x)}} = \rho(y,y',x,t). \tag{A.27}$$

uniformly over $(y,y',x,t) \in \mathcal{W}_y\times\mathcal{W}_y\times B_0\times T_0$. Therefore, as in the proof of (6.35) of Giné et al. (2003), by the bounded convergence theorem, we have

$$\lim_{n\to\infty}\overline{\tau}_{n,0}^2 = \sigma_{0,B}^2.$$

Now, the desired result (A.20) holds if we establish

$$\overline{\tau}_{n,0}^2 - \overline{\sigma}_{n,0}^2 \to 0. \tag{A.28}$$

To show (A.28), set

$$G_n(y,x,y',t) := \sqrt{\rho_2(y,x)\rho_2(y',x+th)}w(y,x)w(y',x+th).$$

Notice that

$$\begin{aligned}(\text{A.29}) \qquad & \int_{T_0}\int_{B_0}\int_{\mathbb{R}}\int_{\mathbb{R}} 1(x\in B_0)1\left(x+th\in B_0\right)G_n(y,y',x,t)dydy'dxdt \\ & \leq \lambda(T_0\times B_0\times \mathcal{W}_y\times \mathcal{W}_y)\sup_{(y,x)\in\mathbb{R}\times B_0}\left|\rho_2(y,x)w^2(y,x)\right| =: \bar{\beta}<\infty.\end{aligned}$$

Let $\varepsilon_n=(\varepsilon_{1n},\varepsilon_{2n})'$,where ε_{1n} and $\varepsilon_{2n}\in(0,h]$ be arbitrary positive sequences such that $\varepsilon_{1n}\to 0$ and $\varepsilon_{2n}/h\to 0$. Define

$$\begin{aligned}\Gamma_1(\varepsilon_n) &:= \left\{(y,y',x,t)\in\mathcal{W}_y\times\mathcal{W}_y\times B_0\times T_0 : \left|y-y'\right|\leq\varepsilon_{1n},\ \|t\|\leq\varepsilon_{2n}/h,\ x+th\in B_0\ \right\}, \\ \Gamma_1^c(\varepsilon_n) &:= \left\{(y,y',x,t)\in\mathcal{W}_y\times\mathcal{W}_y\times B_0\times T_0 : \left|y-y'\right|>\varepsilon_{1n}\text{ or }\varepsilon_{2n}/h<\|t\|\leq 1,\ x+th\in B_0\ \ \right\}.\end{aligned}$$

Let

$$\begin{aligned}\overline{\sigma}^2_{n,0} &= \left(\int\int\int\int_{\Gamma_1(\varepsilon_n)}+\int\int\int\int_{\Gamma_1^c(\varepsilon_n)}\right)Cov\left(\max\{S_{\tau,N}(y,x),0\},\max\{S_{\tau,N}(y',x'),0\}\right) \\ &\quad\times\sqrt{\rho_2(y,x)\rho_2(y',x+th)}w(y,x)w(y',x+th)dydy'dxdt \\ &=: \overline{\sigma}^2_{n,0}(\varepsilon_n)+\overline{\sigma}^2_{n,0,c}(\varepsilon_n)\end{aligned}$$

and

$$\begin{aligned}\overline{\tau}^2_{n,0} &= \left(\int\int\int\int_{\Gamma_1(\varepsilon_n)}+\int\int\int\int_{\Gamma_1^c(\varepsilon_n)}\right)Cov\left(\max\{Z_{1n}(y,x),0\},\max\left\{Z_{2n}(y',x+th),0\right\}\right) \\ &\quad\times\sqrt{\rho_2(y,x)\rho_2(y',x+th)}w(y,x)w(y',x+th)dydy'dxdt \\ &=: \overline{\tau}^2_{n,0}(\varepsilon_n)+\overline{\tau}^2_{n,0,c}(\varepsilon_n).\end{aligned}$$

Then, using (A.25), we have

$$(\text{A.30}) \qquad \overline{\sigma}^2_{n,0}(\varepsilon_n) = \overline{\sigma}^2_{n,0}(0)+O\left(\varepsilon_{1n}\left(\frac{\varepsilon_{2n}}{h}\right)^d\right)=\overline{\sigma}^2_{n,0}(0)+o\left(1\right),$$

$$(\text{A.31}) \qquad \overline{\tau}^2_{n,0}(\varepsilon_n) = \overline{\tau}^2_{n,0}(0)+O\left(\varepsilon_{1n}\left(\frac{\varepsilon_{2n}}{h}\right)^d\right)=\overline{\tau}^2_{n,0}(0)+o(1).$$

Notice that $\overline{\sigma}^2_{n,0}(0)=\overline{\tau}^2_{n,0}(0)$ since $E\max\{S_{\tau,N}(y,x),0\}^2=E\max\{Z_{1n}(y,x),0\}^2=1$. Therefore, to show (A.28), it suffice to establish

$$(\text{A.32}) \qquad \overline{\tau}^2_{n,0,c}(\varepsilon_n)-\overline{\sigma}^2_{n,0,c}(\varepsilon_n)\to 0.$$

Notice that

$$\begin{aligned}&\left|\overline{\tau}^2_{n,0,c}(\varepsilon_n)-\overline{\sigma}^2_{n,0,c}(\varepsilon_n)\right| \\ &=\left|\int\int\int\int_{\Gamma_1^c(\varepsilon_n)}\left[Cov\left(\max\{Z_{1n}(y,x),0\},\max\left\{Z_{2n}(y',x+th),0\right\}\right)\right.\right.\end{aligned}$$

$$-Cov\left(\max\{S_{\tau,N}(y,x),0\},\max\{S_{\tau,N}(y',x+th),0\}\right)\Big]\, G_n(y,y',x,t)dydy'dxdt\Big|$$

$$\leq \int\int\int\int_{\Gamma_1^c(\varepsilon_n)} \big|E\max\{Z_{1n}(y,x),0\}E\max\left\{Z_{2n}(y',x+th),0\right\}$$

$$-E\max\{S_{\tau,N}(y,x),0\}E\max\{S_{\tau,N}(y',x+th),0\}\big|\, G_n(y,y',x,t)dydy'dxdt$$

$$+\int\int\int\int_{\Gamma_1^c(\varepsilon_n)} \big|E\max\{Z_{1n}(y,x),0\}\max\left\{Z_{2n}(y',x+th),0\right\}$$

$$-E\max\{S_{\tau,N}(y,x),0\}\max\{S_{\tau,N}(y',x+th),0\}\big|\, G_n(y,y',x,t)dydy'dxdt$$

$$=: \Delta_{1n}+\Delta_{2n}.$$

We first establish that $\Delta_{1n}=o(1)$ as $n\to\infty$. Let η_1 denote an independent Poisson random variable with mean 1 that is independent of $\{W_i : i\geq 1\}$and set

$$Q_{\tau,n}(y,x)=\left[\sum_{j\leq\eta_1}\varphi(W_j,y,x)-E\varphi(W,y,x)\right]/\sqrt{E\varphi^2(W,y,x)} \tag{A.33}$$

Note that $Var\left(Q_{\tau,n}(y,x)\right)=1$ and for some constant $A_1>0$ independent of $Q_{\tau,n}$ and (y,x),

$$E\left|Q_{\tau,n}(y,x)\right|^3\leq A_1\frac{h^{-3d/2}\left\{E\left|G(Y,y)\phi(x,Z)K\left((x-X)/h\right)\right|^3+E\left|\chi(Z,y,x)K\left((x-X)/h\right)\right|^3\right\}}{\left(h^{-d}E\left[\{G(Y,y)\phi(x,Z)-\chi(Z,y,x)\}K\left((x-X)/h\right)\right]^2\right)^{3/2}}$$

Combining this with Lemma A.5 and Assumption 4.1, we have

$$\sup_{(y,x)\in\mathcal{W}_y\times B_0} E\left|Q_{\tau,n}(y,x)\right|^3=O(h^{-d/2}). \tag{A.34}$$

Let $Q_{\tau,n}^{(1)}(y,x),...,Q_{\tau,n}^{(n)}(y,x)$ be i.i.d. copies of $Q_{\tau,n}(y,x)$. Then obviously, we have

$$\begin{aligned}S_{\tau,N}(y,x)&=\frac{\sqrt{n}\left\{\tau_N(y,x)-E\tau_n(y,x)\right\}}{\sqrt{h^{-2d}E\left[\{G(Y,y)\phi(x,Z)-\chi(Z,y,x)\}K\left((x-X)/h\right)\right]^2}}\\&\overset{d}{=}\frac{\sum_{i=1}^n Q_{\tau,n}^{(i)}(y,x)}{\sqrt{n}}.\end{aligned}$$

Therefore, by (A.7) and (A.34), we have

$$\sup_{(y,x)\in\mathcal{W}_y\times B_0}\left|E\max\{S_{\tau,N}(y,x),0\}-E\max\{Z_{1n}(y,x),0\}\right|\leq O\left(\frac{1}{\sqrt{nh^d}}\right). \tag{A.35}$$

The results (A.29) and (A.35) imply that $\Delta_{1n}=o(1)$ as desired.

We next consider Δ_{2n}. We have

$$\begin{aligned}\Delta_{2n} \leq & \sup_{(y,y',x,t)\in\Gamma_1^c(\varepsilon_n)} \big|E\max\{Z_{1n}(y,x),0\}\max\left\{Z_{2n}(y',x+th),0\right\} \\ & -E\max\{S_{\tau,N}(y,x),0\}\max\{S_{\tau,N}(y',x+th),0\}\big|\cdot\overline{\beta} \\ \text{(A.36)} \qquad & \leq O\left(\varepsilon_{1n}^{-3\alpha_1/2}+\left(\frac{\varepsilon_{2n}}{h}\right)^{-3\alpha_0/2}\right)\cdot O\left(\frac{\log n}{\sqrt{nh^d}}\right) \\ & +O\left(\varepsilon_{1n}^{-3\alpha_1}+\left(\frac{\varepsilon_{2n}}{h}\right)^{-3\alpha_0}\right)\cdot O\left(\frac{(\log n)^2}{nh^{d/2}}\right),\end{aligned}$$

where the first inequality uses (A.29) and the second inequality holds by (A.8) of Lemma A.4, (A.27), (A.34) and Assumption 4.1. Now, since ε_n is arbitrary, we can choose $\varepsilon_{1n}=d_1h^{d/(3\alpha_1)}$ and $\varepsilon_{2n}=d_2h^{1+d/(3\alpha_0)}$ for some constants $d_1>0$ and $d_2>0$. Then, the right hand side of (A.36) is $o(1)$ by our bandwidth condition (Assumption 4.1 (ix)). This establishes (A.32) and hence completes the proof of Lemma A.6. ∎

Let M be defined as in (A.15)-(A.17) and let

$$U_n := \frac{1}{\sqrt{n}}\left\{\sum_{j=1}^{N}1\left((Y_j,X_j)\in\mathbb{R}\times\mathcal{B}(M)\right)-n\Pr\left((Y,X)\in\mathbb{R}\times\mathcal{B}(M)\right)\right\},$$

$$V_n := \frac{1}{\sqrt{n}}\left\{\sum_{j=1}^{N}1\left((Y_j,X_j)\in\mathbb{R}^{d+1}\backslash\left(\mathbb{R}\times\mathcal{B}(M)\right)\right)-n\Pr\left((Y,X)\in\mathbb{R}^{d+1}\backslash\left(\mathbb{R}\times\mathcal{B}(M)\right)\right)\right\}.$$

Also, define

$$S_n := \frac{1}{\sigma_n(B)}T_n^P(B).$$

We next establish the following weak convergence result.

Lemma A.7. *Under Assumption 4.1, we have*

$$(S_n,U_n)\xrightarrow{d}(\mathbb{Z}_1,\sqrt{1-\alpha}\mathbb{Z}_2),$$

where $\mathbb{Z}_1$ and $\mathbb{Z}_2$ are independent $N(0,1)$ random variables and α is defined as in (A.16).

Proof of Lemma A.7. Let

$$\begin{aligned}\text{(A.37)} \qquad \Delta_n(y,x) = & \sqrt{n}\left[\max\{\left[\tau_N(y,x)-E\tau_n(y,x)\right],0\}\right. \\ & \left.-\ E\max\{\left[\tau_N(y,x)-E\tau_n(y,x)\right],0\}\right]w(y,x).\end{aligned}$$

We first construct a partition of $\mathbb{R}\times\mathcal{B}(M)$. Consider the regular grid

$$G_{\mathbf{i}}=(x_{i_1},x_{i_1+1}]\times\cdots\times(x_{i_d},x_{i_d+1}],$$

where $\mathbf{i}=(i_1,...,i_d)$, $i_1,...,i_d$ are integers and $x_i=ih$ for some integer i. Define

$$\begin{aligned}R_{\mathbf{i}} &= \mathbb{R}\times\left(G_{\mathbf{i}}\cap\mathcal{B}(M)\right), \\ \mathcal{I}_n &= \{\mathbf{i}\in\mathbb{Z}^d:(G_{\mathbf{i}}\cap\mathcal{B}(M))\neq\emptyset\}.\end{aligned}$$

Then, we see that $\{R_{\mathbf{i}} : \mathbf{i} \in \mathcal{I}_n \subset \mathbb{Z}^d\}$ is a partition of $\mathbb{R} \times \mathcal{B}(M)$ with

$$\lambda(R_{\mathbf{i}}) \leq A_1 h^d$$
$$m_n := \#(\mathcal{I}_n) \leq A_2 h^{-d}$$

for some positive constants A_1 and A_2, see Mason and Polonik (2009) for a similar construction of partitions in a different context. Letting $u = (y, x)$, set

$$\alpha_{\mathbf{i},n} = \frac{\int_{R_{\mathbf{i}}} 1(u \in B)\Delta_n(u)du}{\sigma_n(B)},$$
$$u_{\mathbf{i},n} = \frac{1}{\sqrt{n}}\left\{\sum_{j=1}^{N} 1\left((Y_j, X_j) \in R_{\mathbf{i}}\right) - n \Pr\left((Y, X) \in R_{\mathbf{i}}\right)\right\}.$$

Then, we have

$$S_n = \sum_{\mathbf{i} \in \mathcal{I}_n} \alpha_{\mathbf{i},n} \text{ and } U_n = \sum_{\mathbf{i} \in \mathcal{I}_n} u_{\mathbf{i},n}.$$

Notice that

$$Var(S_n) = 1 \text{ and } Var(U_n) = 1 - \alpha.$$

For arbitrary $\lambda_1$and $\lambda_2 \in \mathbb{R}$, let

$$y_{\mathbf{i},n} = \lambda_1 \alpha_{\mathbf{i},n} + \lambda_1 u_{\mathbf{i},n}.$$

Notice that $\{y_{\mathbf{i},n} : \mathbf{i} \in \mathcal{I}_n\}$ is an array of mean zero one-dependent random fields. Below we will establish that

$$Var\left(\sum_{\mathbf{i} \in \mathcal{I}_n} y_{\mathbf{i},n}\right) = Var(\lambda_1 S_n + \lambda_2 U_n) \to \lambda_1^2 + \lambda_2^2(1 - \alpha), \tag{A.38}$$

$$\sum_{\mathbf{i} \in \mathcal{I}_n} E\left|y_{\mathbf{i},n}\right|^r = o(1) \text{ for some } 2 < r < 3. \tag{A.39}$$

Then, the result of Lemma A.7 follows from the central limit theorem of Shergin (1990), which is for a triangular array of mean zero m-dependent random fields, and Cramér-Wold device.

We first establish (A.38), which holds if we have

$$Cov\left(S_n, U_n\right) = O\left(\frac{1}{\sqrt{nh^{2d}}}\right). \tag{A.40}$$

Now, (A.40) holds if

$$Cov\left(\int_{\mathbb{R}}\int_{B_0} \sqrt{n} \max\{\left[\tau_N(y, x) - E\tau_n(y, x)\right], 0\} w(y, x) dx dy, U_n\right) = O\left(\frac{1}{\sqrt{nh^{2d}}}\right). \tag{A.41}$$

For any $(y, x) \in \mathcal{W}_y \times B_0$, we have

$$\left(S_{\tau,N}(y, x), \frac{U_n}{\sqrt{\Pr\left((Y, X) \in \mathbb{R} \times \mathcal{B}(M)\right)}}\right) \stackrel{d}{=} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left(Q_{\tau,n}^{(i)}(y, x), U^{(i)}\right), \tag{A.42}$$

where $\left(Q_{\tau,n}^{(i)}(y,x), U^{(i)}\right)$ for $i = 1, ..., n$ are i.i.d. copies of $(Q_{\tau,n}(y,x), U)$, with $Q_{\tau,n}(y,x)$ defined as in (A.33) and

$$U = \left[\sum_{j \le \eta_1} 1\left((Y_j, X_j) \in \mathbb{R} \times \mathcal{B}(M)\right) - \Pr\left((Y,X) \in \mathbb{R} \times \mathcal{B}(M)\right)\right] / \sqrt{\Pr\left((Y,X) \in \mathbb{R} \times \mathcal{B}(M)\right)}.$$

Let $(Z_{1n}(y,x), Z_{2n})$ for $(y,x) \in \mathbb{R}^{d+1}$, be a mean zero multivariate Gaussian process such that, for each $(y,x) \in \mathbb{R}^{d+1}$, $(Z_{1n}(y,x), Z_{2n})$ and the left-hand side of (A.42) have the same covariance structure. That is,

$$\begin{aligned}&(Z_{1n}(y,x), Z_{2n})\\ &\overset{d}{=} \left(\sqrt{1-(\gamma_n^*(y,x))^2}\mathbb{Z}_1 + \gamma_n^*(y,x)\mathbb{Z}_2, \mathbb{Z}_2\right),\end{aligned}$$

where $\mathbb{Z}_1$ and $\mathbb{Z}_2$ are independent standard normal random variables and

$$\gamma_n^*(y,x) = E\left[S_{\tau,N}(y,x)\frac{U_n}{\sqrt{\Pr\left((Y,X) \in \mathbb{R} \times \mathcal{B}(M)\right)}}\right].$$

Notice that we have

$$\sup_{\mathcal{W}_y \times B_0} |\gamma_n^*(y,x)| = O(h^{d/2}), \tag{A.43}$$

which in turn is less than or equal to ε for all sufficiently large n and any $0 < \varepsilon < 1/2$. This result and (A.9) imply that

$$\begin{aligned}&\sup_{\mathcal{W}_y \times B_0}\left|Cov\left(\max\{S_{\tau,N}(y,x), 0\}, \frac{U_n}{\sqrt{\Pr\left((Y,X) \in \mathbb{R} \times \mathcal{B}(M)\right)}}\right) - E\max\{Z_{1n}(y,x),0\}Z_{2n}\right|\\ &\le O\left(\frac{1}{\sqrt{nh^{2d}}}\right).\end{aligned} \tag{A.44}$$

On the other hand,

$$\begin{aligned}\sup_{\mathcal{W}_y \times B_0} |E\max\{Z_{1n}(y,x),0\}Z_{2n}| &= \sup_{\mathcal{W}_y \times B_0} |\gamma_n^*(y,x)E\max\{Z_{1n}(y,x),0\}Z_{1n}(y,x)|\\ &\le \sup_{\mathcal{W}_y \times B_0} |\gamma_n^*(y,x)|\, EZ_{1n}^2(y,x)\\ &= \sup_{\mathcal{W}_y \times B_0} |\gamma_n^*(y,x)| = O(h^{d/2}),\end{aligned} \tag{A.45}$$

using the law of iterated expectations and (A.43). Therefore, (A.44) and (A.45) imply that

$$\sup_{\mathcal{W}_y \times B_0}\left|Cov\left(\sqrt{n}\max\{[\tau_N(y,x) - E\tau_n(y,x)], 0\}, U_n\right)\right| \le O\left(\frac{1}{\sqrt{nh^{2d}}} + h^{d/2}\right),$$

which, when combined with $\lambda(\mathcal{W}_y \times B_0) < \infty$, yields (A.41) and hence (A.38), as desired.

We next establish (A.39). Notice that, with $2 < r < 3$,

$$\sigma_n^r(B)E\left|\alpha_{\mathbf{i},n}\right|^r$$

$$\text{(A.46)} \qquad \leq \left(\int_{R_{\mathbf{i}}} \int_{R_{\mathbf{i}}} \int_{R_{\mathbf{i}}} 1_B(u) 1_B(v) 1_B(w) E \left| \Delta_n(u) \Delta_n(v) \Delta_n(w) \right| dudvdw \right)^{r/3},$$

where $1_B(u) = 1(u \in B)$ by the Liapunov inequality. Also, using Jensen's inequality and the elementary result $|\max\{X, 0\}| \leq |X|$, we have:

$$\text{(A.47)} \qquad E \left| \Delta_n(y, x) \right|^3 \leq 8n^{3/2} E \left| \tau_N(y, x) - E\tau_n(y, x) \right|^3.$$

By Rosenthal's inequality (see, e.g., Lemma 2.3 of Giné et al., 2003), we have:

$$\text{(A.48)} \qquad \sup_{\mathcal{W}_y \times B_0} n^{3/2} E \left| \tau_N(y, x) - E\tau_n(y, x) \right|^3 \leq O\left(\frac{1}{h^{3d/2}} + \frac{1}{n^{1/2} h^{2d}} \right).$$

Now, (A.46), (A.47), (A.48), the elementary result $E|XYZ| \leq E(|X| + |Y| + |Z|)^3$ and the facts that $\lambda(R_{\mathbf{i}}) \leq A_1 h^d$, $nh^d \to \infty$, and $\sigma_n^r(B) = O(1)$ imply that

$$\text{(A.49)} \qquad E \left| \alpha_{\mathbf{i},n} \right|^r \leq O(h^{rd/2}) \text{ uniformly in } \mathbf{i} \in \mathcal{I}_n.$$

Therefore, we have

$$\text{(A.50)} \qquad \sum_{\mathbf{i} \in \mathcal{I}_n} E \left| \alpha_{\mathbf{i},n} \right|^r \leq O(m_n h^{rd/2}) = O(h^{(r/2-1)d}) = o(1).$$

On the other hand, set

$$p_{\mathbf{i},n} = \Pr\left((Y, X) \in R_{\mathbf{i}} \right).$$

Then, by the Rosenthal's inequality, there exists a constant $D_1 > 0$ such that

$$\text{(A.51)} \qquad \begin{aligned} \sum_{\mathbf{i} \in \mathcal{I}_n} E \left| u_{\mathbf{i},n} \right|^r &\leq D_1 n^{-r/2} \sum_{\mathbf{i} \in \mathcal{I}_n} \left((np_{\mathbf{i},n})^{r/2} + np_{\mathbf{i},n} \right) \\ &\leq D_1 \max_{\mathbf{i} \in \mathcal{I}_n} \left((p_{\mathbf{i},n})^{(r-2)/2} + n^{-1/2} \right) \to 0. \end{aligned}$$

Therefore, combining (A.50) and (A.51), we have (A.39). This now completes the proof of Lemma A.7. ∎

Lemma A.8. *Under Assumption 4.1, we have*

$$\begin{aligned} \lim_{n \to \infty} \int_{B_0} \int_{\mathbb{R}} \left[\sqrt{n} E \max\{ [\tau_N(y, x) - E\tau_n(y, x)], 0\} - E \max\{\mathbb{Z}, 0\} k_{\tau,n}^{1/2}(y, x) \right] w(y, x) dy dx &= 0, \\ \lim_{n \to \infty} \int_{B_0} \int_{\mathbb{R}} \left[\sqrt{n} E \max\{ [\tau_n(y, x) - E\tau_n(y, x)], 0\} - E \max\{\mathbb{Z}, 0\} k_{\tau,n}^{1/2}(y, x) \right] w(y, x) dy dx &= 0, \end{aligned}$$

where $\mathbb{Z}$ is a standard normal random variable.

Proof of Lemma A.8. This result follows from Lemma A.4 and an argument similar to proof of Lemma 6.3 of Giné et al. (2003). ∎

Let

$$\text{(A.52)} \qquad \begin{aligned} L_n(B) = \frac{\sqrt{n}}{\sigma_n(B)} \int_{B_0} \int_{\mathbb{R}} &[\max\{ [\tau_n(y, x) - E\tau_n(y, x)], 0\} \\ &- E \max\{ [\tau_n(y, x) - E\tau_n(y, x)], 0\}] w(y, x) dy dx. \end{aligned}$$

Lemma A.9. *Under Assumption 4.1, we have*

$$L_n(B) \xrightarrow{d} \mathbb{Z}$$

as $n \to \infty$, where $\mathbb{Z}$ stands for the standard normal random variable.

Proof of Lemma A.9. Notice that

$$\begin{aligned} S_n &= \frac{\sqrt{n}}{\sigma_n(B)} \int_{B_0} \int_{\mathbb{R}} [\max\{[\tau_N(y,x) - E\tau_n(y,x)], 0\} \\ &\quad -E\max\{[\tau_N(y,x) - E\tau_n(y,x)], 0\}] \, w(y,x) dy dx. \end{aligned}$$

Conditional on $N = n$, we have

$$\begin{aligned} S_n &\overset{d}{=} \frac{\sqrt{n}}{\sigma_n(B)} \int_{B_0} \int_{\mathbb{R}} [\max\{[\tau_n(y,x) - E\tau_n(y,x)], 0\} \\ &\quad -E\max\{[\tau_n(y,x) - E\tau_n(y,x)], 0\}] \, w(y,x) dy dx. \end{aligned} \tag{A.53}$$

By Lemma A.6 and the de-Poissonization argument of Beirlant and Mason (1995) (see also Lemma 2.4 of Giné et al., 2003), we have

$$\begin{aligned} &\frac{\sqrt{n}}{\sigma_n(B)} \int_{B_0} \int_{\mathbb{R}} [\max\{[\tau_n(y,x) - E\tau_n(y,x)], 0\} \\ &-E\max\{[\tau_N(y,x) - E\tau_n(y,x)], 0\}] \, w(y,x) dy dx \xrightarrow{d} \mathbb{Z}. \end{aligned}$$

Now the result of Lemma A.9 follows from Lemma A.8, which implies

$$\begin{aligned} &\lim_{n\to\infty} \int_{B_0} \int_{\mathbb{R}} \left[\sqrt{n} E \max\{[\tau_N(y,x) - E\tau_n(y,x)], 0\}\right. \\ &\left.-\sqrt{n} E \max\{[\tau_n(y,x) - E\tau_n(y,x)], 0\}\right] w(y,x) dy dx = 0. \quad \blacksquare \end{aligned}$$

Lemma A.10. *Let $\{W_j \in \mathbb{R}^k : j = 1, ..., n\}$ be i.i.d random vectors with $E\|W\| < \infty$. Let $h : \mathbb{R}^k \times \mathbb{R}^k \to \mathbb{R}$ be a real function such that $Eh(W, w) = 0$ for all $w \in \mathbb{R}^k$. Let*

$$T_n = \int_{\mathcal{B}} \max\left\{ \sum_{j=1}^n h(W_j, w), 0 \right\} dw,$$

where $\mathcal{B} \subset \mathbb{R}^k$ is a Borel set. Then, for any convex function $g : \mathbb{R} \to \mathbb{R}$, we have

$$Eg(T_n - ET_n) \le Eg\left(2\sum_{j=1}^n \varepsilon_j \int_{\mathcal{B}} |h(W_j, w)| \, dw \right),$$

where $\{\varepsilon_j : j = 1, ..., n\}$ are i.i.d random variables with $\Pr(\varepsilon = 1) = \Pr(\varepsilon = -1) = 1/2$, *independent of $\{W_i : i = 1, ..., n\}$.*

Proof of Lemma A.10. We can establish Lemma A.10 by modifying the majorization inequality results of Pinelis (1994). Let $(W_1^*, ..., W_n^*)$ be an independent copy of $(W_1, ..., W_n)$. For $i = 1, ..., n$, let E_i and E_i^* denote the conditional expectations given $(W_1, ..., W_i)$ and $(W_1, ..., W_{i-1}, W_i^*)$. Let

$$\xi_i = E_i T_n - E_{i-1} T_n, \tag{A.54}$$

$$\eta_i = E_i \left(T_n - T_{n,-i}\right), \tag{A.55}$$

where

$$T_{n,-i} = \int_{\mathcal{B}} \max\left\{\sum_{j\neq i}^{n} h(W_j, w), 0\right\} dw,$$

Then, we have

$$T_n - ET_n = \xi_1 + \cdots \xi_n, \tag{A.56}$$

$$\xi_i = \eta_i - E_{i-1}\eta_i, \tag{A.57}$$

$$|\eta_i| \leq \int_{\mathcal{B}} |h(X_i, w)|\, dw, \tag{A.58}$$

where (A.56) follows from (A.54), (A.57) holds by independence of W_j 's, and (A.58) follows from the elementary inequality $|\max\{a+b, 0\} - \max\{a, 0\}| \leq |b|$. Let

$$\eta_i^* = E_i^* \left(T_{n,i}^* - T_{n,-i}\right),$$

where

$$T_{n,i}^* = \int_{\mathcal{B}} \max\left\{\sum_{j\neq i}^{n} h(W_j, w) + h(W_i^*, w), 0\right\} dw.$$

Notice that the random variables η_i and η_i^* are conditionally independent given $(W_1, ..., W_{i-1})$, and the conditional distributions of η_i and η_i^* given $(W_1, ..., W_{i-1})$ are equivalent. Therefore, for any convex function $f : \mathbb{R} \to \mathbb{R}$, we have

$$\begin{aligned}
E_{i-1} f(\xi_i) &= E_{i-1} f(\eta_i - E_{i-1}\eta_i) \\
&\leq E_{i-1} f(\eta_i - E_{i-1}\eta_i - \eta_i^* - E_{i-1}\eta_i^*) \\
&\leq \frac{1}{2} E_{i-1}\left[f(2\eta_i) + f(-2\eta_i^*)\right] \\
&\leq \frac{1}{2} E_{i-1}\left[f\left(2\int_{\mathcal{B}} |h(X_i, w)|\, dw\right) + f\left(-2\int_{\mathcal{B}} |h(X_i, w)|\, dw\right)\right] \\
&= Ef\left(2\varepsilon_i \int_{\mathcal{B}} |h(X_i, w)|\, dw\right),
\end{aligned}$$

where the first inequality follows from Berger (1991, Lemma 2.2), the second inequality holds by the convexity of f and the last inequality follows from the convexity of f and (A.58). Now, the result of Lemma A.10 holds by (A.56) and Lemma 2.6 of Berger (1991). ∎

Lemma A.11. *Let Assumption 4.1 hold. Then, for any Borel subset $A_0 \subset \mathbb{R}^d$ and $A_1 \subset \mathbb{R}$, we have*

$$\overline{\lim_{n\to\infty}} E\left(\sqrt{n}\int_{A_0}\int_{A_1} \{h_n(y, x) - Eh_n(y, x)\}\, w(y, x) dy dx\right)^2$$

$$\leq \; C_0 \int_{A_0}\int_{A_0}\int_{A_1}\int_{A_1} f(x)g(x,y,y')w(y,x)w(y',x')dydy'dxdx'$$

for some constant $C_0 > 0$, *where*

$$\begin{aligned} h_n(y,x) &= \max\{\tau_n(y,x) - E\tau_n(y,x), 0\}, \\ g(x,y,y') &= g_1(x,y,y') + g_2(x,y,y'), \\ g_1(z,y,y') &= E\left[\left|G(Y,y)\phi(x,Z)G(Y,y')\phi(x,Z)\right| |X=z\right], \\ g_2(z,y,y') &= E\left[\left|\chi(Z,y,x)\chi(Z,y',x)\right| |X=z\right]. \end{aligned}$$

Proof of Lemma A.11. We have

$$\begin{aligned}
&E\left(\sqrt{n}\int_{A_0}\int_{A_1}\{h_n(y,x) - Eh_n(y,x)\}\, w(y,x)dydx\right)^2 \\
&\leq 8E\left(\frac{1}{h^d}\int_{A_0}\int_{A_1}\left|K\left(\frac{x-X}{h}\right)G(Y,y)\phi(x,Z)\right| w(y,x)dydx\right)^2 \\
&+ 8E\left(\frac{1}{h^d}\int_{A_0}\int_{A_1}\left|K\left(\frac{x-X}{h}\right)\chi(Z,y,x)\right| w(y,x)dydx\right)^2 \\
&\leq 8\left(\sup_u |K(u)|\right) \\
&\times\left\{E\int_{A_0}\int_{A_0}\int_{A_1}\int_{A_1}\left|\frac{1}{h^d}K\left(\frac{x-X}{h}\right)\right|\left|G(Y,y)\phi(x,Z)G(Y,y')\phi(x',Z)\right| w(y,x)w(y',x')dydy'dxdx'\right. \\
&\left.+ E\int_{A_0}\int_{A_0}\int_{A_1}\int_{A_1}\left|\frac{1}{h^d}K\left(\frac{x-X}{h}\right)\right|\left|\chi(Z,y,x)\chi(Z,y',x')\right| w(y,x)w(y',x')dydy'dxdx'\right\} \\
&\leq 8\left(\sup_u |K(u)|\right)^2 \\
&\times\left\{\int_{A_0}\int_{A_0}\int_{A_1}\int_{A_1}\frac{1}{h^d}E\left[1\left(X\in\left[x-\frac{h}{2}, x+\frac{h}{2}\right]^d\right)g(X,y,y')\right]w(y,x)w(y',x')dydy'dxdx'\right\} \\
&\leq 8\left(\sup_u |K(u)|\right)^2\int_{A_0}\int_{A_0}\int_{A_1}\int_{A_1} f(x)g(x,y,y')w(y,x)w(y',x')dydy'dxdx' + o(1)
\end{aligned}$$

as $n \to \infty$, where the first inequality follows from Lemma A.10 and the inequality $(a+b)^2 \leq 2(a^2+b^2)$ and the last convergence holds by bounded convergence theorem. Now, take $C_0 = 8\left(\sup_u |K(u)|\right)^2$ to get the desired result of Lemma A.11. ∎

We are now ready to prove Theorem A.2.

Proof of Theorem A.2. Let $\{B_{0k} : k \geq 1\}$ be a sequence of Borel sets in $\mathbb{R}^d$ that has finite Lebesgue measure $\lambda(B_{0k}) < \infty$ and satisfies (A.15)-(A.17) with $B_0 = B_{0k}$ for each k and

$$\lim_{k\to\infty}\int_{B_{0k}^c\cap\mathcal{W}_x} f(x)dx = 0. \tag{A.59}$$

Let $B_k = \mathbb{R} \times B_{0k}$. Then, for each $k \geq 1$, by Lemma A.9, we have

$$L_n(B_k) \xrightarrow{d} \mathbb{Z}$$

and, by Lemma A.6,

$$\lim_{n\to\infty} \sigma_n^2(B_k) = \sigma_{0,B_k}^2.$$

By Lemma A.11, we have

(A.60)

$$\begin{aligned}
&\overline{\lim_{n\to\infty}} E\left(\sqrt{n}\int_{B_{0k}^c}\int_{\mathbb{R}} \{\max\{\tau_n(y,x) - E\tau_n(y,x), 0\} - E\max\{\tau_n(y,x) - E\tau_n(y,x), 0\}\}\, w(y,x)dydx\right)^2 \\
&\leq C_1 \int_{B_{0k}^c\cap\mathcal{W}_x}\int_{B_{0k}^c\cap\mathcal{W}_x}\int_{\mathbb{R}}\int_{\mathbb{R}} f(x)g(x,y,y')w(y,x)w(y',x')dydy'dxdx' \\
&\leq C_2\lambda(\mathcal{W}_y\times\mathcal{W}_y\times\mathcal{W}_x)\left(\sup_{(x,y,y',x)\in\mathcal{W}_x\times\mathcal{W}_y\times\mathcal{W}_y} g(x,y,y')\right)\int_{B_{0k}^c\cap\mathcal{W}_x} f(x)dx,
\end{aligned}$$

where C_1 and C_2 are positive constants. Also,

$$\lim_{k\to\infty} \sigma_{0,B_k}^2 = \sigma_0^2. \tag{A.61}$$

Therefore, by (A.59), (A.60), (A.61) and Theorem 4.2 of Billingsley (1968), we conclude that

$$\begin{aligned}
&\int\int \sqrt{n}\left[\max\{[\tau_n(y,x) - E\tau_n(y,x)]\,,0\} - E\max\{[\tau_n(y,x) - E\tau_n(y,x)]\,,0\}\right] w(y,x)dydx \\
&\xrightarrow{d} \sigma_0\mathbb{Z}.
\end{aligned}$$

Now, the proof of Theorem A.2 is complete since, using Lemma A.8, we have

$$\lim_{n\to\infty}\left|\int\int \sqrt{n}E\max\{[\tau_n(y,x) - E\tau_n(y,x)]\,,0\}w(y,x)dydx - a_n\right| = 0. \quad \blacksquare$$

A.4. **The Asymptotic Power Properties.**

Proof of Theorem 4.2. Using Lemma A.7 and Assumption 4.1 (iii), we have

$$\begin{aligned}
&\left|n^{-1/2}\hat{T} - \int\int \max\{\tau_0(y,x), 0\}\, w(y,x)dydx\right| \\
&\leq \int\int |\hat{\tau}(y,x) - \tau_0(y,x)|\, w(y,x)dydx \\
&\xrightarrow{p} 0.
\end{aligned} \tag{A.62}$$

Therefore,

$$\begin{aligned}
\Pr\left(\hat{S} > z_{1-\alpha}\right) &= \Pr\left(\hat{T} > \hat{a}_n + \hat{\sigma}z_{1-\alpha}\right) \\
&= \Pr\left(n^{-1/2}\hat{T} > n^{-1/2}\left(\hat{a}_n + \hat{\sigma}z_{1-\alpha}\right)\right) \\
&= \Pr\left(n^{-1/2}\hat{T} > 0\right) + o(1) \\
&\to 1,
\end{aligned}$$

where the third equality holds by Theorem A.2 and Assumption 4.1 (ix) which implies $(nh^d)^{-1/2} \to 0$ and the last convergence to one follows from (A.62) and the definition of the alternative hypothesis $H_1 : \int\int \max\{\tau_0(y,x), 0\} w(y,x)dydx > 0$. ∎

Proof of Theorem 4.3. Under $H_a : \tau_0(y,x) = n^{-1/2}\delta(y,x)$, we will show below that

$$\frac{T_n^* - \tilde{a}_n}{\sigma_0} \xrightarrow{d} N(0,1), \tag{A.63}$$

where σ_0 is defined in the main text and

$$\tilde{a}_n = \int\int E\max\left\{\delta(y,x) + h^{-d/2}\sqrt{\rho_2(y,x)}\mathbb{Z}, 0\right\} w(y,x)dydx.$$

Notice that

(A.64)

$$\begin{aligned}
&\tilde{a}_n - a_n \\
&= \int\int E\left[\max\left\{\delta(y,x) + h^{-d/2}\sqrt{\rho_2(y,x)}\mathbb{Z}, 0\right\} - \max\left\{h^{-d/2}\sqrt{\rho_2(y,x)}\mathbb{Z}, 0\right\}\right] w(y,x)dydx \\
&\geq \frac{1}{2}\int\int \delta(y,x)w(y,x)dydx > 0,
\end{aligned}$$

where the inequality holds by the general result that $a \geq [\max\{a+b, 0\} - \max\{b, 0\}] \geq a1(b \geq 0)$ for $a > 0$ and $E1(\mathbb{Z} \geq 0) = 1/2$. Therefore, we have

$$\begin{aligned}
\Pr\left(\hat{S} > z_{1-\alpha}\right) &= \Pr\left(\hat{T} > \hat{a}_n + \hat{\sigma}z_{1-\alpha}\right) \\
&= \Pr\left(T_n^* > \hat{a}_n + \hat{\sigma}z_{1-\alpha}\right) + o(1) \\
&= \Pr\left(\frac{T_n^* - \tilde{a}_n}{\sigma_0} > \frac{\hat{a}_n - a_n}{\sigma_0} + \frac{\hat{\sigma}}{\sigma_0}z_{1-\alpha} - \frac{\tilde{a}_n - \hat{a}_n}{\sigma_0}\right) \\
&> \Pr\left(\frac{T_n^* - \tilde{a}_n}{\sigma_0} > \frac{\hat{a}_n - a_n}{\sigma_0} + \frac{\hat{\sigma}}{\sigma_0}z_{1-\alpha}\right) \\
&= 1 - \Phi\left(z_{1-\alpha}\right) + o(1) \to \alpha,
\end{aligned}$$

where the second equality holds by Lemma A.6, the third equality holds by rearranging terms, the inequality holds by (A.64), and the last equality follows from (A.63) and Theorem A.1, as desired.

It now suffices to establish (A.63). Its proof is similar to that of Theorem A.2 and we briefly sketch the main difference. Notice first that Lemmas A.1-A.3 hold under H_a without any modification. Define B as before and now set a poissonization version of T_n^* (restricted to B) to be:

$$\begin{aligned}
T_n^{*P}(B) &= \int_{B_0}\int_{\mathbb{R}} \max\{\delta(y,x) + \sqrt{n}\left[\tau_N(y,x) - E\tau_n(y,x)\right], 0\}w(y,x)dydx \\
&\quad - \int_{B_0}\int_{\mathbb{R}} E\max\{\delta(y,x) + \sqrt{n}\left[\tau_N(y,x) - E\tau_n(y,x)\right], 0\}w(y,x)dydx. \qquad \text{(A.65)}
\end{aligned}$$

Using (A.23) - (A.25), we have that

$$Var\left(T_n^{*P}(B)\right)$$

$$
\begin{aligned}
&= \int_{B_0}\int_{B_0}\int_{\mathbb{R}}\int_{\mathbb{R}} 1\left(\|x-x'\|\le h\right) \\
&\quad \times Cov\left(\max\left\{\frac{\delta(y,x)}{\sqrt{k_{\tau,n}(y,x)}}+S_{\tau,N}(y,x),0\right\},\max\left\{\frac{\delta(y',x')}{\sqrt{k_{\tau,n}(y',x')}}+S_{\tau,N}(y',x'),0\right\}\right) \\
&\quad \times\sqrt{k_{\tau,n}(y,x)k_{\tau,n}(y',x')}w(y,x)w(y',x')dydy'dxdx' \\
&= \overline{\sigma}^2_{n,0}+o(1),
\end{aligned}
$$

where $\overline{\sigma}^2_{n,0}$ is defined as in (A.26). Then, using the same arguments as in Lemma A.6, we can see that

$$
\lim_{n\to\infty} Var\left(T_n^{*P}(B)\right)=\lim_{n\to\infty}\sigma_n^2(B)=\sigma^2_{0,B}, \tag{A.66}
$$

where $\sigma^2_{0,B}$ is defined in (A.21). Lemma A.7 also holds under H_a with S_n and $\Delta_n(y,x)$ now defined by

$$
S_n=\frac{1}{\sigma_n(B)}T_n^{*P}(B)
$$

and

$$
\begin{aligned}
\Delta_n(y,x) &= \Big[\max\left\{\delta(y,x)+\sqrt{n}\left[\tau_N(y,x)-E\tau_n(y,x)\right],0\right\} \\
&\qquad E\max\left\{\delta(y,x)+\sqrt{n}\left[\tau_N(y,x)-E\tau_n(y,x)\right],0\right\}\Big]\,w(y,x),
\end{aligned}
$$

respectively, and by applying the CLT for 1-independent triangular arrays. On the other hand, Lemma A.8 should be modified to:

$$
\begin{aligned}
&\lim_{n\to\infty}\int_{B_0}\int_{\mathbb{R}}\Bigg[E\max\Big\{\delta(y,x)+\sqrt{n}\big[\tau_N(y,x)-E\tau_n(y,x)\big],0\Big\} \\
&-E\max\Big\{\delta(y,x)+h^{-d/2}\sqrt{\rho_2(y,x)}\mathbb{Z},0\Big\}\Bigg]w(y,x)dydx=0, \\
&\lim_{n\to\infty}\int_{B_0}\int_{\mathbb{R}}\Bigg[E\max\{\delta(y,x)+\sqrt{n}\left[\tau_n(y,x)-E\tau_n(y,x)\right],0\} \\
&-E\max\Big\{\delta(y,x)+h^{-d/2}\sqrt{\rho_2(y,x)}\mathbb{Z},0\Big\}\Bigg]w(y,x)dydx=0.
\end{aligned}
$$

Also, Lemma A.9 should hold with $L_n(B)$ now defined by

$$
\begin{aligned}
L_n(B) &= \frac{1}{\sigma_n(B)}\int_{B_0}\int_{\mathbb{R}}\big[\max\{\delta(y,x)+\sqrt{n}\left[\tau_n(y,x)-E\tau_n(y,x)\right],0\} \\
&\qquad -E\max\{\delta(y,x)+\sqrt{n}\left[\tau_n(y,x)-E\tau_n(y,x)\right],0\}\big]\,w(y,x)dydx.
\end{aligned}
$$

The remaining proof of (A.63) is analogous to the proof of Theorem A.2. This completes the proof of Theorem 4.3. ∎

A.5. **Proofs for Section 5.1.** For $r>0$, define the r-enlargement of the contact set C to be

$$
C(r)=\{(y,x)\in\mathcal{W}:|\tau_0(y,x)|\le r\}.
$$

Also, let

$$\begin{aligned}
\underline{\mathbf{1}}_n &= 1\left(\int\int_{C((1-\varepsilon)\eta_n+\epsilon)} w(y,x)dydx > 0\right), \\
\overline{\mathbf{1}}_n &= 1\left(\int\int_{C((1+\varepsilon)\eta_n+\epsilon)} w(y,x)dydx > 0\right), \\
\hat{\mathbf{1}}_n &= 1\left(\int\int_{\hat{C}_\epsilon} w(y,x)dydx > 0\right),
\end{aligned}$$

where $1(\cdot)$ denotes the indicator function. The following lemmas are useful to prove Theorem 5.1.

Lemma A.12. *Under Assumption 4.1, for each $\varepsilon > 0$,*

$$\Pr\left\{C\left((1-\varepsilon)\eta_n+\epsilon\right) \subset \hat{C}_\epsilon \subset C\left((1+\varepsilon)\eta_n+\epsilon\right)\right\} \to 1.$$

Proof of Lemma A.12. Using Lemma A.3, we have

$$\Pr\{\sup_{(y,x)\in\mathcal{W}} |\hat{\tau}(y,x) - \tau_0(y,x)| > \varepsilon\eta_n\} \to 0 \tag{A.67}$$

by the choice η_n which satisfies $(nh^d)^{1/2}\eta_n/\log n \to \infty$ and $h^{-s}\eta_n \to \infty$ by Assumptions 4.1 (ix) and 5.1 (ii). Therefore, (A.67) implies that, for any $(y,x) \in C\left((1-\varepsilon)\eta_n+\epsilon\right)$, by the triangle inequality,

$$|\hat{\tau}(y,x)| \leq (1-\varepsilon)\eta_n + \epsilon + |\hat{\tau}(y,x) - \tau_0(y,x)| \leq \eta_n + \epsilon,$$

with probability approaching one. Thus we deduce that $\Pr\left\{C\left((1-\varepsilon)\eta_n+\epsilon\right) \subset \hat{C}_\epsilon\right\} \to 1$. Now, for any $(y,x) \in \hat{C}_\epsilon$, by the triangular inequality,

$$|\tau_0(y,x)| \leq \eta_n + \epsilon + |\hat{\tau}(y,x) - \tau_0(y,x)| \leq (1+\varepsilon)\eta_n + \epsilon,$$

with probability approaching one. Therefore, $\Pr\left\{\hat{C}_\epsilon \subset C\left((1+\varepsilon)\eta_n+\epsilon\right)\right\} \to 1$, as desired. ∎

Lemma A.13. *Suppose Assumptions 4.1 and 5.1 hold. Then, we have*

$$\begin{aligned}
(a)\ \hat{a}_n(\hat{C}_\epsilon) &= a_n(C_\epsilon) + o_p(1), \\
(b)\ \hat{\sigma}^2(\hat{C}_\epsilon) &= \sigma_0^2(C_\epsilon) + o_p(1).
\end{aligned}$$

Proof of Lemma A.13. Let $u = (y,x)$ and, for a Borel set $A \subset \mathbb{R}^{d+1}$, define $\lambda_\rho(\cdot)$ to be

$$\lambda_\rho(A) = \int_A \sqrt{\rho_2(u)}w(u)du.$$

By the triangle inequality, we have

$$\begin{aligned}
&\left|\int_{\hat{C}_\epsilon} \sqrt{\hat{\rho}_2(u)}w(u)du - \int_{C_\epsilon} \sqrt{\rho_2(u)}w(u)du\right| \\
&\leq \int_{\hat{C}_\epsilon \Delta C_\epsilon} \sqrt{\rho_2(u)}w(u)du + \int_{\hat{C}_\epsilon} \left|\sqrt{\hat{\rho}_2(u)} - \sqrt{\rho_2(u)}\right| w(u)du \\
&=: D_{1n} + D_{2n},
\end{aligned}$$

where Δ denotes the symmetric difference.

Let

$$\tilde{C}_\epsilon = \{u \in \mathcal{W} : |\tau_0(u)| \leq 2\eta_n + \epsilon\},$$
$$E_n = \{u \in \mathcal{W} : |\hat{\tau}(u) - \tau_0(u)| > \eta_n\}.$$

We first establish $D_{1n} = o_p(h^{d/2})$ by extending the result of Cuevas and Fraiman (1997, Theorem 1). We have

$$\begin{aligned}
D_{1n} &= \lambda_\rho(\hat{C}_\epsilon \Delta C_\epsilon) = \lambda_\rho(\hat{C}_\epsilon \cap C_\epsilon^c) + \lambda_\rho(\hat{C}_\epsilon^c \cap C_\epsilon) \\
&\leq \lambda_\rho(\hat{C}_\epsilon \cap \tilde{C}_\epsilon^c) + \lambda_\rho(\tilde{C}_\epsilon \cap C_\epsilon^c) + \lambda_\rho(\hat{C}_\epsilon^c \cap C_\epsilon) \\
&= \lambda_\rho(\hat{C}_\epsilon \cap \tilde{C}_\epsilon^c \cap E_n) + \lambda_\rho(\tilde{C}_\epsilon \cap C_\epsilon^c) + \lambda_\rho(\hat{C}_\epsilon^c \cap C_\epsilon \cap E_n) \\
&\leq 2\lambda_\rho(E_n) + b_n,
\end{aligned} \tag{A.68}$$

where $b_n = h^*(2\eta_n)$, the first inequality uses $C_\epsilon \subset \tilde{C}_\epsilon$, the third equality follows from the facts that $\lambda_\rho(\hat{C}_\epsilon \cap \tilde{C}_\epsilon^c \cap E_n^c) = 0$ and $\lambda_\rho(\hat{C}_\epsilon^c \cap C_\epsilon \cap E_n^c) = 0$ and the last inequality holds by $\lambda_\rho(A) \leq \lambda_\rho(B)$ for $A \subset B$ and Assumption *5.1* (ii).

Now, using Lemma A.3, we have that

$$\int |\hat{\tau}(u) - \tau_0(u)| \sqrt{\rho_2(u)} w(u) du = O_p\left[\left(nh^d\right)^{-1/2} \log n + h^s\right]. \tag{A.69}$$

Let $\rho_n = \min\{\left(nh^d\right)^{1/2} (\log n)^{-1}, h^{-s}\}$. Then, for any $\varepsilon > 0$,

$$\begin{aligned}
\Pr\left(h^{-d/2} D_{1n} > \varepsilon\right) &\leq \Pr\left(2\lambda_\rho(E_n) + b_n > \varepsilon h^{d/2}\right) \\
&\leq \Pr\left(\frac{1}{\eta_n} \int |\hat{\tau}(u) - \tau_0(u)| \sqrt{\rho_2(u)} w(u) du > \frac{\varepsilon h^{d/2} - b_n}{2}\right) \\
&\leq \Pr\left(\rho_n \int |\hat{\tau}(u) - \tau_0(u)| \sqrt{\rho_2(u)} w(u) du > \frac{\varepsilon \rho_n \eta_n h^{d/2}}{2}\right) + o(1) \\
&\to 0,
\end{aligned} \tag{A.70}$$

where the first inequality holds by (A.68), the second inequality holds by the inequality $1(E_n) \leq |\hat{\tau}(u) - \tau_0(u)| / \eta_n$, the third inequality follows from Assumption *5.1* (ii) which implies $\rho_n \eta_n b_n \to 0$, and the last convergence to zero holds by (A.69) and $\rho_n \eta_n h^{d/2} \to \infty$. This now establishes that $D_{1n} = o_p(h^{d/2})$.

We next consider D_{2n}. We have

$$\begin{aligned}
h^{-d/2} D_{2n} &\leq \left[\inf_{u \in \mathcal{W}} |\rho_2(u)| + o_p(1)\right]^{-1/2} h^{-d/2} \int |\hat{\rho}_2(u) - \rho_2(u)| w(u) du \\
&= O_p(n^{-1/2} h^{-d} \log n + h^{s-d/2}) \to 0,
\end{aligned} \tag{A.71}$$

where the inequality holds with probability that goes to 1 using Lemma A.3, the equality holds by Lemma A.3 and the last convergence to zero holds by our assumption on h. (A.69) and (A.71) now establish part (a) of Theorem B.2. The proof of part (b) is similar. ∎

Lemma A.14. *Suppose Assumptions 4.1 and 5.1 hold and* $\int\int_{C_\epsilon} w(y,x)dydx > 0$. *Then, under the null hypothesis* H_0, *we have*

$$T_n^* = \tilde{T}_n(C_\epsilon) + o_p(1), \tag{A.72}$$

where T_n^* *is defined in (A.3) and*

$$\tilde{T}_n(C_\epsilon) = \int\int_{C_\epsilon} \sqrt{n}\max\{\tau_0(y,x) + [\tau_n(y,x) - E\tau_n(y,x)], 0\}w(y,x)dydx. \tag{A.73}$$

Proof of Lemma A.14. Under the null hypothesis, we have $\tau_0(y,x) \le 0$ for all $(y,x) \in \mathcal{Y}\times\mathcal{X}$. Therefore, for each $\varepsilon > 0$, we can write

$$\begin{aligned}
&\left|T_n^* - \tilde{T}_n(C_\epsilon)\right| \\
&= \int\int 1\left((y,x)\in A_{1\epsilon}(\varepsilon)\right)\sqrt{n}\max\{\tau_0(y,x) + [\tau_n(y,x) - E\tau_n(y,x)], 0\}w(y,x)dydx \\
&+ \int\int 1\left((y,x)\in A_{2\epsilon}(\varepsilon)\right)\sqrt{n}\max\{\tau_0(y,x) + [\tau_n(y,x) - E\tau_n(y,x)], 0\}w(y,x)dydx \\
&=: D_{1n} + D_{2n},
\end{aligned} \tag{A.74}$$

where

$$\begin{aligned}
A_{1\epsilon}(\varepsilon) &= \{(y,x)\in\mathcal{W} : -\varepsilon-\epsilon \le \tau_0(y,x) < -\epsilon\}, \\
A_{2\epsilon}(\varepsilon) &= \{(y,x)\in\mathcal{W} : \tau_0(y,x) < -\varepsilon-\epsilon\}.
\end{aligned}$$

Choose $\varepsilon = \varepsilon_n$ such that $h^{-d/2}(\log n)\varepsilon_n^\gamma \to 0$ and $(nh^d)^{1/2}(\log n)^{-1}\varepsilon_n \to \infty$. Then, we have

$$\begin{aligned}
D_{1n} &\le \int\int 1\left((y,x)\in A_{1\epsilon}(\varepsilon_n)\right)\sqrt{n}\max\{[\tau_n(y,x) - E\tau_n(y,x)], 0\}w(y,x)dydx \\
&\le \sup|w(y,x)|\cdot\sqrt{n}\sup_{(y,x)\in\mathcal{W}}|\tau_n(y,x) - E\tau_n(y,x)|\cdot\lambda\left(A_{1\epsilon}(\varepsilon_n)\right) \\
&\le O_p\left[h^{-d/2}\log n\right]\cdot O(\varepsilon_n^\gamma) = o_p(1),
\end{aligned} \tag{A.75}$$

where the last inequality holds by (A.5) and Assumption *5.1* (ii). Also, we have

$$\begin{aligned}
\Pr(D_{2n} > \delta) &\le \Pr\left(\sup_{(y,x)\in\mathcal{W}}\{[\tau_n(y,x) - E\tau_n(y,x)] - \varepsilon_n - \epsilon\} > 0\right) \\
&\le \Pr\left(\sup_{(y,x)\in\mathcal{W}}|\tau_n(y,x) - E\tau_n(y,x)| > \varepsilon_n + \epsilon\right) \to 0,
\end{aligned} \tag{A.76}$$

where the convergence to zero holds by (A.5) and by the choice of ε_n. This establishes Lemma A.14. ∎

Define

$$T_n(C_\epsilon) = \int\int_{C_\epsilon} \sqrt{n}\max\{[\tau_n(y,x) - E\tau_n(y,x)], 0\}w(y,x)dydx. \tag{A.77}$$

Lemma A.15. *Suppose Assumptions 4.1 and 5.1 hold and* $\int\int_{C_\epsilon} w(y,x)dydx > 0$*. Then, we have*

$$\frac{T_n(C_\epsilon) - a_n(C_\epsilon)}{\sigma_0(C_\epsilon)} \Rightarrow N(0,1).$$

Proof of Lemma A.15. The proof of Lemma B.4 is similar to that of Theorem A.2. ∎

Proof of Theorem 5.1. Consider part (a) first. Write

$$\hat{S}^* = \hat{S}_C \hat{\mathbf{1}}_n + \hat{S}\left(1 - \hat{\mathbf{1}}_n\right)$$

By Lemma A.12, we know that

$$\underline{\mathbf{1}}_n \le \hat{\mathbf{1}}_n \le \overline{\mathbf{1}}_n \text{ wp} \to 1.$$

First, suppose that $\int\int_{C_\epsilon} w(y,x)dydx > 0$. Then, $\underline{\mathbf{1}}_n = 1$ and hence $\hat{\mathbf{1}}_n = 1$ wp$\to$ 1. Therefore, we have

$$\begin{aligned}
\Pr\left(\hat{S}^* > z_{1-\alpha}\right) &= \Pr\left(\hat{S}_C > z_{1-\alpha}\right) + o(1) \\
&= \Pr\left(\frac{\tilde{T}_n(C_\epsilon) - a_n(C_\epsilon)}{\sigma_0(C_\epsilon)} > z_{1-\alpha}\right) + o(1) \\
&\le \Pr\left(\frac{T_n(C_\epsilon) - a_n(C_\epsilon)}{\sigma_0(C_\epsilon)} > z_{1-\alpha}\right) + o(1) \\
&= \alpha + o(1),
\end{aligned} \tag{A.78}$$

where the second equality follows from Lemmas A.2, A.13 and A.14, the inequality is due to the fact that $T_n(C_\epsilon) \ge \tilde{T}_n(C_\epsilon)$, and the last equality holds by Lemma A.15.

Next, suppose that $\int\int_C w(y,x)dydx = 0$. Then, $\overline{\mathbf{1}}_n = 0$ for n sufficiently large and hence $\hat{\mathbf{1}}_n = 0$ wp$\to$ 1. Therefore,

$$\begin{aligned}
\Pr\left(\hat{S}^* > z_{1-\alpha}\right) &= \Pr\left(\hat{S} > z_{1-\alpha}\right) + o(1) \\
&\le \alpha + o(1),
\end{aligned} \tag{A.79}$$

by Theorem 4.1 (a). Now, (A.78) and (A.79) combine to yield the desired result.

Part (b) can be proven in the same manner as the proof of Theorem 4.2 (a), except that we now have that

$$\begin{aligned}
\Pr\left(\hat{S}^* > z_{1-\alpha}\right) &= \Pr\left(\hat{T} > \left(\hat{a}_n(\hat{C}_\epsilon) + \hat{\sigma}(\hat{C}_\epsilon)z_{1-\alpha}\right)\hat{\mathbf{1}}_n + (\hat{a}_n + \hat{\sigma}z_{1-\alpha})\left(1 - \hat{\mathbf{1}}_n\right)\right) \\
&= \Pr\left(n^{-1/2}\hat{T} > 0\right) + o(1) \to 1,
\end{aligned}$$

since $n^{-1/2}\hat{T}$ converges in probability to a positive constant under H_1. ∎

Proof of Theorem 5.2. We first establish part (a). Using Lemma A.3 and Assumption 5.1 (ii), under H_a^*, we have

$$\begin{aligned} &\Pr\{\sup_{(y,x)\in\mathcal{W}} |\hat{\tau}(y,x) - \mu(y,x)| > \varepsilon\eta_n\} \\ \leq\ &\Pr\{\sup_{(y,x)\in\mathcal{W}} |\hat{\tau}(y,x) - \tau_0(y,x)| > \varepsilon\eta_n - n^{-1/2} \sup_{(y,x)\in\mathcal{W}} |\delta(y,x)|\} \\ \to\ &0, \end{aligned} \tag{A.80}$$

since $n^{-1/2}\sup_{(y,x)\in\mathcal{W}} |\delta(y,x)|$ can be made arbitrarily small for n sufficiently large using Assumption 5.2 (iii). Let

$$C_a(r) = \{(y,x) \in \mathcal{W} : |\mu(y,x)| \leq r\}.$$

Then, under H_a^*, we have that for each $\varepsilon > 0$,

$$\Pr\left\{C_a\left((1-\varepsilon)\,\eta_n/2 + \epsilon\right) \subset \hat{C}_\epsilon\right\} \to 1 \tag{A.81}$$

because for any $(y,x) \in C\left((1-\varepsilon)\,\eta_n/2 + \epsilon\right)$,

$$|\hat{\tau}(y,x)| \leq (1-\varepsilon)\eta_n + \epsilon + |\hat{\tau}(y,x) - \tau_0(y,x)| \leq \eta_n + \epsilon,$$

wp $\to 1$. Since $0 < \int\int_{C_a} w(y,x)dydx \leq \int\int_{C_a((1-\varepsilon)\eta_n/2+\epsilon)} w(y,x)dydx$, (A.81) implies that $\hat{\mathbf{1}}_n = 1$ wp $\to 1$. Therefore, it follows that,

$$\hat{S}^* = \hat{S}_C \ \text{ wp} \to 1 \text{ under } H_a. \tag{A.82}$$

Below, we shall establish that

$$\hat{T} = T_n^*(C_a) + o_p(1), \tag{A.83}$$

$$\hat{a}_n(\hat{C}_\epsilon) = a_n(C_a) + o_p(1), \tag{A.84}$$

$$\hat{\sigma}^2(\hat{C}_\epsilon) = \sigma_0^2(C_a) + o_p(1), \tag{A.85}$$

and

$$\frac{T_n^*(C_a) - \tilde{a}_n(C_a)}{\sigma_0(C_a)} \xrightarrow{d} N(0,1), \tag{A.86}$$

where

$$\begin{aligned} T_n^*(C_a) &= \int\int_{C_a} \max\{\delta(y,x) + \sqrt{n}\left[\tau_n(y,x) - E\tau_n(y,x)\right], 0\} w(y,x)dydx. \\ \tilde{a}_n(C_a) &= \int\int_{C_a} E\max\left\{\delta(y,x) + h^{-d/2}\sqrt{\rho_2(y,x)}\mathbb{Z}, 0\right\} w(y,x)dydx. \end{aligned}$$

Now, we can get the desired result of part (a) because

$$\begin{aligned} \Pr\left(\hat{S}^* > z_{1-\alpha}\right) &= \Pr\left(\hat{S}_C > z_{1-\alpha}\right) + o(1) \\ &= \Pr\left(\hat{T} > \hat{a}_n(\hat{C}_\epsilon) + \ \hat{\sigma}(\hat{C}_\epsilon)z_{1-\alpha}\right) + o(1) \\ &= \Pr\left(T_n^*(C_a) > a_n(C_a) + \ \sigma_0(C_a)z_{1-\alpha}\right) + o(1) \end{aligned} \tag{A.87}$$

$$\begin{aligned}
&= \Pr\left(\frac{T_n^*(C_a)-\tilde{a}_n(C_a)}{\sigma_0(C_a)} > \frac{a_n(C_a)-\tilde{a}_n(C_a)}{\sigma_0(C_a)} + \ z_{1-\alpha}\right)+o(1)\\
&> \Pr\left(\frac{T_n^*(C_a)-\tilde{a}_n(C_a)}{\sigma_0(C_a)} > \ z_{1-\alpha}\right)+o(1)\\
&\to \alpha,
\end{aligned}$$

where the first equality holds by (A.82), the second equality follows from the definition of $\hat{S}_C$, the third equality holds by (A.83), (A.84) and (A.85), the last convergence to α holds by (A.86) and the inequality holds because

(A.88)

$$\begin{aligned}
&\tilde{a}_n(C_a)-a_n(C_a)\\
&=\int\int_{C_a} E\left[\max\left\{\delta(y,x)+h^{-d/2}\sqrt{\rho_2(y,x)}\mathbb{Z},0\right\}-\max\left\{h^{-d/2}\sqrt{\rho_2(y,x)}\mathbb{Z},0\right\}\right]w(y,x)dydx\\
&\geq \frac{1}{2}\int\int_{C_a}\delta(y,x)w(y,x)dydx>0.
\end{aligned}$$

It remains to establish (A.83) - (A.86). First, (A.83) follows by Lemma A.2 and a modification of the proof of Lemma A.14. In particular, C_ϵ is replaced by C_a, $\tau_0(y,x)$ in $A_{1\epsilon}(\varepsilon)$ and $A_{2\epsilon}(\varepsilon)$ is replaced by $\mu(y,x)$. Then (A.75) is replaced by

$$\begin{aligned}
D_{1n} &\leq \int\int 1\left((y,x)\in A_{1\epsilon}(\varepsilon)\right)\max\{\delta(y,x)+\sqrt{n}\left[\tau_n(y,x)-E\tau_n(y,x)\right],0\}w(y,x)dydx\\
&\leq \sup|w(y,x)|\cdot\left\{\sup_{(y,x)\in\mathcal{W}}|\delta(y,x)|+\sqrt{n}\sup_{(y,x)\in\mathcal{W}}|\tau_n(y,x)-E\tau_n(y,x)|\right\}\cdot\lambda\left(A_{1\epsilon}(\varepsilon)\right)\\
&= o_p(1),
\end{aligned}$$

using Assumption 5.2 (iv), and the inequality (A.76) is replaced by

$$\begin{aligned}
\Pr\left(D_{2n}>\delta\right) &\leq \Pr\left(\sup_{(y,x)\in\mathcal{W}}\left\{n^{-1/2}\delta(y,x)+\left[\tau_n(y,x)-E\tau_n(y,x)\right]-\varepsilon-\epsilon\right\}>0\right)\\
&\leq \Pr\left(\sup_{(y,x)\in\mathcal{W}}|\tau_n(y,x)-E\tau_n(y,x)|>\varepsilon_0+\epsilon\right),
\end{aligned}$$

for some $\varepsilon_0>0$ and n sufficiently large, using the assumption $\sup_{(y,x)\in\mathcal{W}}|\delta(y,x)|<\infty$. Second, (A.84) also holds by a modification of the proof of Lemma A.13. That is, $\tilde{C}_\epsilon$ is now defined with $\tau(u)$ replace by $\mu(u)=\mu(y,x)$, E_n is defined with $\hat{\tau}(u)-\tau_0(u)$ replaced by $\hat{\tau}(u)-\mu(u)$, b_n is defined as $b_n=h^{**}(2\eta_n)$, and the inequality (A.70) is replace by

$$\begin{aligned}
\Pr\left(h^{-d/2}D_{1n}>\varepsilon\right) &\leq \Pr\left(2\lambda_\rho(E_n)+b_n>\varepsilon h^{d/2}\right)\\
&\leq \Pr\left(\frac{1}{\eta_n}\int|\hat{\tau}(u)-\mu(u)|\sqrt{\rho_2(u)}w(u)du>\frac{\varepsilon h^{d/2}-b_n}{2}\right)\\
&\leq \Pr\left(\rho_n\int\left\{|\hat{\tau}(u)-\tau_0(u)|+n^{-1/2}|\delta(u)|\right\}\sqrt{\rho_2(u)}w(u)du>\frac{\varepsilon\rho_n\eta_n h^{d/2}}{2}\right)+o(1)
\end{aligned}$$

$$\to \quad 0$$

using the condition $\rho_n n^{-1/2} \to 0$. (A.85) can be verified in a similar fashion. Finally, (A.86) can be verified following the same steps using those in the proof of Theorem A.2. This establishes part (a) of Theorem 5.2.

We next consider part (b). Let

$$\varepsilon = \sigma_0^{-1}(C_a) \int\int_{C_a} \delta(y,x) w(y,x) dydx > 0.$$

Note that

$$0 < \frac{1}{2}\varepsilon \le d_{1n} := \frac{\tilde{a}_n(C_a) - a_n(C_a)}{\sigma_0(C_a)} \le \varepsilon < \infty, \tag{A.89}$$

For n sufficiently large, we have

(A.90)

$$d_{2n} := \frac{a_n - a_n(C_a)}{\sigma_0(C_a)} = h^{-d/2} \int\int_{\mathbb{R}^2 \setminus C_a} \sqrt{\rho_2(y,x)} w(y,x) dydx \cdot E \max\{\mathbb{Z}_1, 0\} \sigma_0^{-1}(C_a) > 2\varepsilon > 0.$$

Let

$$Q_n = \frac{T_n^*(C_a) - \tilde{a}_n(C_a)}{\sigma_0(C_a)}.$$

Then, we have

$$\begin{aligned}
& \lim_{n\to\infty} \left[\Pr\left(\hat{S}^* > z_{1-\alpha}\right) - \Pr\left(\hat{S} > z_{1-\alpha}\right) \right] \\
= & \lim_{n\to\infty} \left[\Pr\left(Q_n > -d_{1n} + z_{1-\alpha}\right) - \Pr\left(Q_n > d_{2n} - d_{1n} + \frac{\sigma_0}{\sigma_0(C_a)} z_{1-\alpha}\right) \right] \\
\ge & \lim_{n\to\infty} \left[\Pr\left(Q_n > -d_{1n} + z_{1-\alpha}\right) - \Pr\left(Q_n > d_{2n} - d_{1n} + z_{1-\alpha}\right) \right] \\
> & \lim_{n\to\infty} \left[\Pr\left(Q_n > -\frac{\varepsilon}{2} + z_{1-\alpha}\right) - \Pr\left(Q_n > 2\varepsilon - \varepsilon + z_{1-\alpha}\right) \right] \\
= & \lim_{n\to\infty} \left[\Pr\left(Q_n \in [-\varepsilon + z_{1-\alpha}, -\frac{\varepsilon}{2} + z_{1-\alpha})\right) \right] \\
= & \Pr\left(\mathbb{Z} \in [-\varepsilon + z_{1-\alpha}, -\frac{\varepsilon}{2} + z_{1-\alpha})\right) > 0,
\end{aligned}$$

where the first equality holds by Theorem A.1, (A.83), and (A.87), the first inequality uses the fact $\sigma_0/\sigma_0(C_a) \ge 1$, the second inequality holds by (A.89) and (A.90) and the last equality holds by (A.86). This completes the proof of Theorem 5.2. ∎

Table 1. Nonparametric tests using data from LaLonde (1986)

	outcome (Y): RE78			outcome (Y): RE78-RE75		
	p-values			P-values		
Bandwidth	weight function			weight function		
(h)	Uniform	Inverse S.E.	Density	Uniform	Inverse S.E.	Density
$H_0 : E[Y_1 - Y_0 \mid X = x] = 0$ for each $x \in \mathcal{W}_x$ vs.						
$H_1 : E[Y_1 - Y_0 \mid X = x] > 0$ for some $x \in \mathcal{W}_x$						
2.021	0.042	0.019	0.021	0.003	0.003	0.026
2.526	0.032	0.021	0.030	0.002	0.005	0.036
3.031	0.032	0.029	0.050	0.003	0.009	0.058
3.537	0.032	0.037	0.073	0.004	0.017	0.089
$H_0 : E[Y_1 - Y_0 \mid X = x] = 0$ for each $x \in \mathcal{W}_x$ vs.						
$H_1 : E[Y_1 - Y_0 \mid X = x] \neq 0$ for some $x \in \mathcal{W}_x$						
5.052	0.233	0.329	0.452	0.078	0.267	0.492
6.063	0.198	0.277	0.395	0.089	0.271	0.489
7.073	0.177	0.240	0.347	0.079	0.226	0.423
8.084	0.162	0.211	0.306	0.073	0.200	0.379
$H_0 : E[1(Y_1 \leq y) \mid X = x] \leq E[1(Y_0 \leq y) \mid X = x]$ for each $(y, x) \in \mathcal{W}$ vs.						
$H_1 : E[1(Y_1 \leq y) \mid X = x] > E[1(Y_0 \leq y) \mid X = x]$ for some $(y, x) \in \mathcal{W}$						
2.021	0.840	0.851	0.846	0.817	0.697	0.699
2.526	0.930	0.910	0.849	0.855	0.781	0.642
3.031	0.948	0.932	0.870	0.903	0.850	0.668
3.537	0.937	0.925	0.860	0.891	0.837	0.658
$H_0 : E[1(Y_1 \leq y) \mid X = x] = E[1(Y_0 \leq y) \mid X = x]$ for each $(y, x) \in \mathcal{W}$ vs.						
$H_1 : E[1(Y_1 \leq y) \mid X = x] \neq E[1(Y_0 \leq y) \mid X = x]$ for some $(y, x) \in \mathcal{W}$						
5.052	0.150	0.623	0.124	0.032	0.145	0.117
6.063	0.148	0.601	0.120	0.034	0.140	0.123
7.073	0.153	0.592	0.128	0.037	0.147	0.130
8.084	0.154	0.572	0.140	0.042	0.147	0.144

Note: The table shows p-values for four different combinations of null and alternative hypotheses. Three types of weight functions were used: the uniform weight, the inverse standard-error weight, and the density weight.

Table 2. Results of Monte Carlo experiments [one-sided test].

Sample Size (n)	Bandwidth (h)	DGP1: H_0 is true (least favorable case) Nominal Probabilities 0.10	0.05	0.01	DGP2: Mimicking the NSW data Nominal Probabilities 0.10	0.05	0.01
Uniform weight $w_1(x) \equiv 1$							
722	2.019	0.113	0.064	0.021	0.848	0.763	0.547
	2.524	0.109	0.064	0.021	0.836	0.750	0.537
	3.028	0.103	0.060	0.021	0.806	0.716	0.506
	3.533	0.103	0.060	0.021	0.780	0.683	0.481
1444	1.656	0.120	0.068	0.021	0.976	0.967	0.910
	2.070	0.106	0.062	0.018	0.987	0.972	0.900
	2.484	0.108	0.063	0.019	0.980	0.958	0.878
	2.898	0.107	0.064	0.020	0.971	0.945	0.851
Inverse-standard-error weight $\hat{w}_2(x) = [\hat{\rho}_2(x)]^{-1/2}$							
722	2.019	0.107	0.062	0.021	0.887	0.821	0.649
	2.524	0.101	0.058	0.019	0.865	0.779	0.588
	3.028	0.097	0.058	0.018	0.815	0.722	0.519
	3.533	0.096	0.058	0.018	0.771	0.670	0.471
1444	1.656	0.109	0.060	0.018	0.981	0.979	0.960
	2.070	0.102	0.054	0.016	0.996	0.989	0.952
	2.484	0.101	0.054	0.017	0.991	0.977	0.921
	2.898	0.100	0.055	0.017	0.980	0.958	0.880
Density weight $\hat{w}_3(x) = \hat{p}_1(x) \cdot \hat{p}_0(x)$							
722	2.019	0.109	0.062	0.019	0.850	0.772	0.575
	2.524	0.107	0.065	0.021	0.816	0.717	0.517
	3.028	0.106	0.064	0.021	0.741	0.636	0.436
	3.533	0.106	0.064	0.022	0.679	0.574	0.379
1444	1.656	0.104	0.063	0.018	0.979	0.971	0.931
	2.070	0.106	0.064	0.019	0.991	0.976	0.913
	2.484	0.105	0.065	0.020	0.979	0.955	0.868
	2.898	0.104	0.063	0.019	0.956	0.916	0.794

Note: The table shows coverage probabilities of testing the null hypothesis of zero CATE for every $x \in \mathcal{W}_x$ against the alternative hypothesis of positive CATE for some $x \in \mathcal{W}_x$ (one-sided test). Three types of weight functions were used: the uniform weight, the inverse standard-error weight, and the density weight.

Table 3. Results of Monte Carlo experiments [two-sided test].

Sample Size (n)	Bandwidth (h)	DGP1: H_0 is true (least favorable case) Nominal Probabilities 0.10	0.05	0.01	DGP2: Mimicking the NSW data Nominal Probabilities 0.10	0.05	0.01
Uniform weight $w_1(x) \equiv 1$							
722	5.047	0.109	0.066	0.021	0.520	0.416	0.245
	6.056	0.106	0.064	0.021	0.510	0.407	0.240
	7.066	0.105	0.062	0.020	0.501	0.399	0.233
	8.075	0.102	0.059	0.020	0.493	0.388	0.229
1444	4.140	0.107	0.061	0.018	0.826	0.747	0.577
	4.967	0.102	0.060	0.018	0.814	0.737	0.567
	5.795	0.101	0.059	0.018	0.807	0.729	0.554
	6.623	0.101	0.058	0.020	0.796	0.718	0.541
Inverse-standard-error weight $\hat{w}_2(x) = [\hat{\rho}_2(x)]^{-1/2}$							
722	5.047	0.113	0.066	0.023	0.486	0.380	0.218
	6.056	0.110	0.063	0.022	0.471	0.372	0.210
	7.066	0.108	0.064	0.021	0.462	0.364	0.206
	8.075	0.109	0.064	0.021	0.455	0.358	0.211
1444	4.140	0.110	0.063	0.018	0.814	0.732	0.544
	4.967	0.108	0.062	0.018	0.784	0.696	0.507
	5.795	0.106	0.061	0.019	0.773	0.679	0.494
	6.623	0.105	0.060	0.019	0.757	0.668	0.482
Density weight $\hat{w}_3(x) = \hat{p}_1(x) \cdot \hat{p}_0(x)$							
722	5.047	0.109	0.063	0.022	0.403	0.309	0.163
	6.056	0.110	0.065	0.022	0.393	0.300	0.160
	7.066	0.109	0.064	0.024	0.388	0.298	0.166
	8.075	0.110	0.065	0.024	0.386	0.299	0.171
1444	4.140	0.107	0.062	0.021	0.698	0.596	0.409
	4.967	0.108	0.064	0.021	0.655	0.552	0.370
	5.795	0.107	0.065	0.023	0.641	0.539	0.359
	6.623	0.109	0.066	0.024	0.628	0.528	0.354

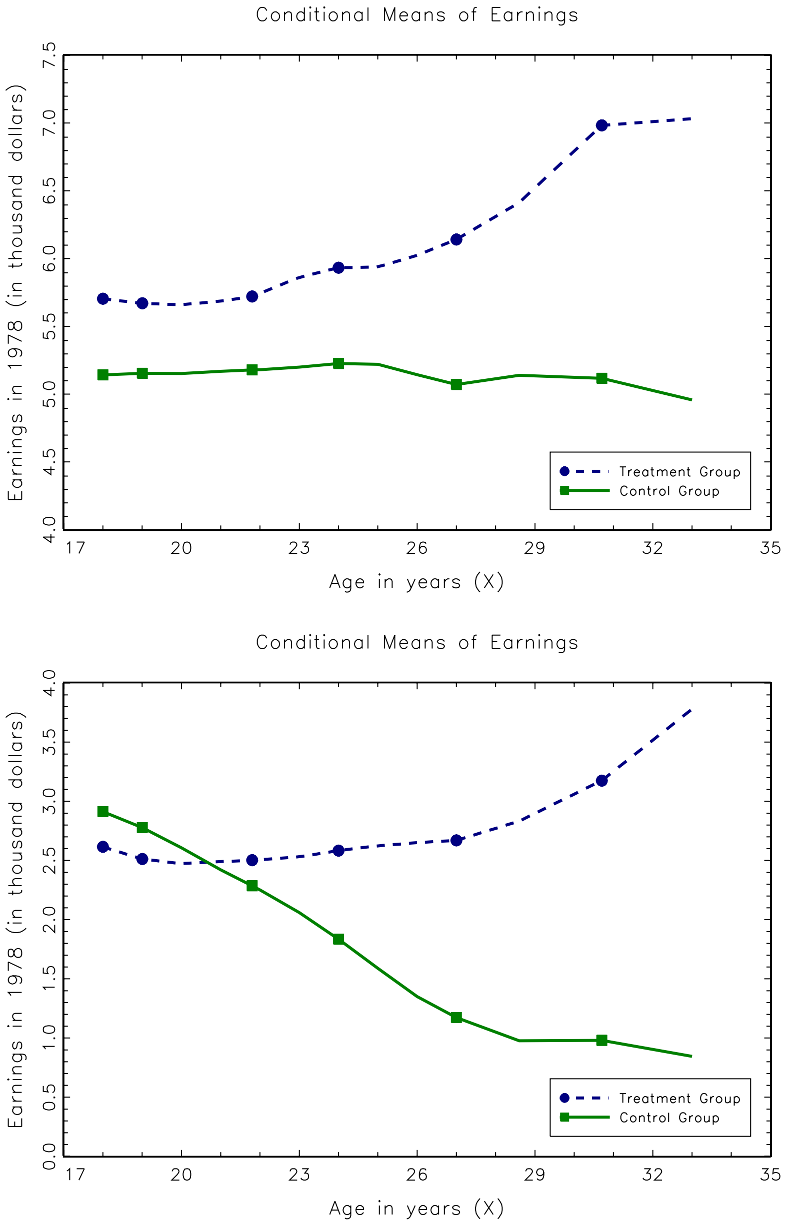
Note: The table shows coverage probabilities of testing the null hypothesis of zero CATE for every $x \in \mathcal{W}_x$ against the alternative hypothesis of nonzero CATE for some $x \in \mathcal{W}_x$ (two-sided test). Three types of weight functions were used: the uniform weight, the inverse standard-error weight, and the density weight.

TABLE 4. Results of Monte Carlo experiments: Coverage Probabilities of Nonparametric Tests of Crump et al. (2008)

Sample Size (n)	Order of Power Series ($K-1$)	DGP1: H_0 is true (least favorable case) Nominal Probabilities 0.10	0.05	0.01	DGP2: Mimicking the NSW data Nominal Probabilities 0.10	0.05	0.01
H_0: the CATE is zero for each x (Test Statistic T)							
722	1	0.099	0.072	0.028	0.512	0.432	0.293
	2	0.108	0.068	0.022	0.452	0.368	0.236
	3	0.064	0.045	0.023	0.163	0.127	0.069
	4	0.063	0.039	0.018	0.136	0.096	0.056
	5	0.065	0.040	0.020	0.459	0.403	0.301
1444	1	0.114	0.080	0.042	0.798	0.741	0.626
	2	0.114	0.075	0.040	0.740	0.653	0.516
	3	0.071	0.051	0.026	0.218	0.173	0.100
	4	0.049	0.032	0.016	0.169	0.120	0.051
	5	0.051	0.028	0.014	0.725	0.671	0.573
H_0: the CATE is zero for each x (Test Statistic Q)							
722	1	0.098	0.042	0.004	0.504	0.362	0.140
	2	0.100	0.045	0.007	0.443	0.295	0.113
	3	0.059	0.034	0.013	0.153	0.097	0.033
	4	0.059	0.030	0.010	0.131	0.074	0.029
	5	0.061	0.031	0.011	0.453	0.355	0.231
1444	1	0.109	0.059	0.013	0.794	0.686	0.409
	2	0.105	0.055	0.012	0.725	0.587	0.336
	3	0.065	0.036	0.010	0.211	0.138	0.049
	4	0.045	0.021	0.004	0.166	0.086	0.024
	5	0.046	0.020	0.004	0.714	0.626	0.462

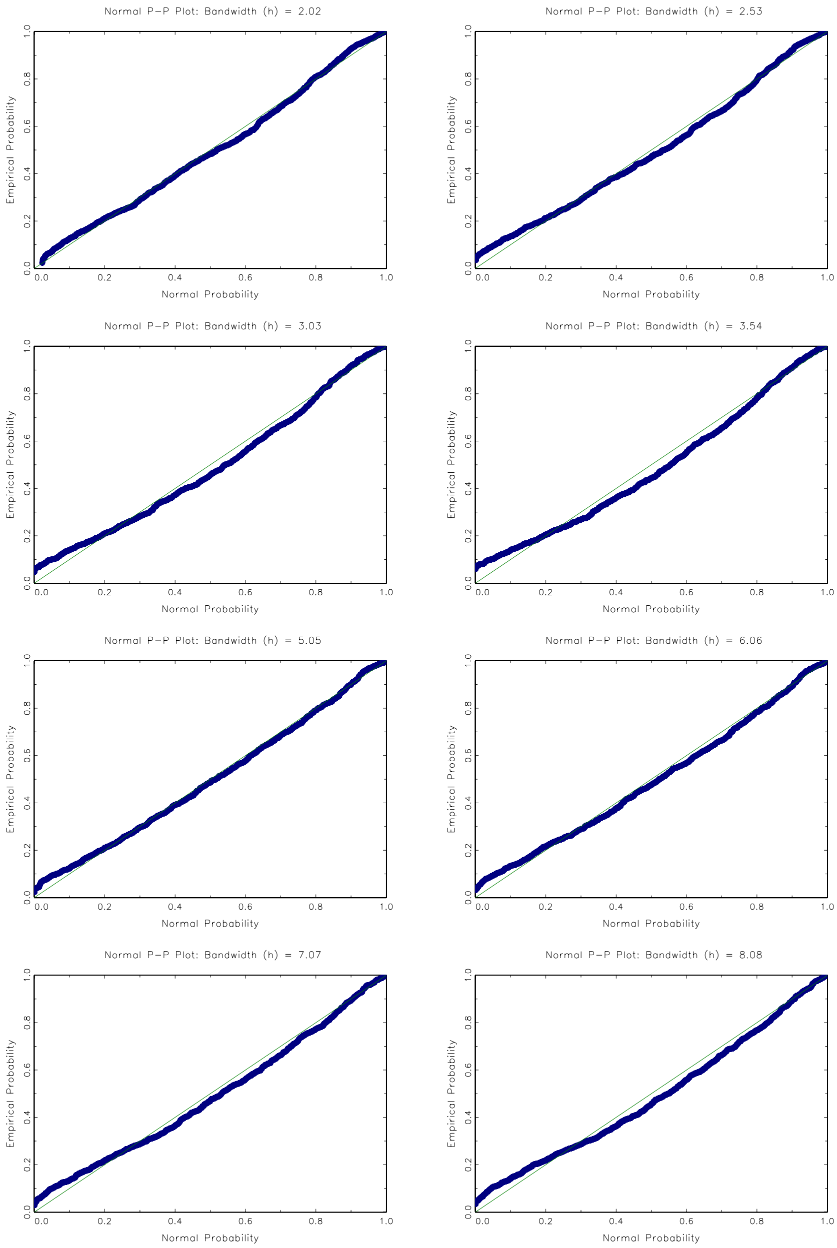
Note: The top panel of the table shows coverage probabilities of the nonparametric test of Crump et al. (2008) with their statistic T for the null hypothesis that the conditional average treatment effect (CATE) is zero for each value of x. The bottom panel shows coverage probabilities of the nonparametric test of Crump et al. (2008) with their statistic Q.

FIGURE 1. Nonparametric estimation of conditional treatment effects



Note: The top panel shows nonparametric estimates of conditional means of earnings in 1978 (Y, in thousand dollars) as functions of age in years (X) for the treatment and control groups, respectively. The bottom panel shows nonparametric estimates of conditional means of changes in earnings between 1978 and 1975.

FIGURE 2. Results of Monte Carlo experiments: normal P-P plots



Note: The top four figures show normal P-P plots for the one-sided nonparametric test of the null hypothesis that the conditional average treatment effect (CATE) is negative for each value of x. Each panel of the figure shows a P-P plot with a different value of the bandwidth (h). These figures corresponds to the results reported in the top panel of Table 2 with $n = 722$. The bottom four figures show normal P-P plots for the two-sided test. These figures correspond to the results reported in the bottom panel of Table 3 with $n = 722$.

References

Abadie, A. (2002). Bootstrap tests for distributional treatment effects in instrumental variables models. *Journal of the American Statistical Association 97*(457), 284–292.

Abadie, A., J. Angrist, and G. Imbens (2002). Instrumental variables estimates of the effect of subsidized training on the quantiles of trainee earnings. *Econometrica 70*(1), 91–117.

Abadie, A. and G. W. Imbens (2006). Large sample properties of matching estimators for average treatment effects. *Econometrica 74*(1), 235–267.

Abbring, J. H. and J. J. Heckman (2007). Econometric evaluation of social programs, part III: Distributional treatment effects, dynamic treatment effects, dynamic discrete choice, and general equilibrium polic evaluation. In J. J. Heckman and E. E. Leamer (Eds.), *Handbook of Econometrics*, Volume 6, Chapter 72, pp. 5145–5303. Elsevier.

Anderson, G., O. Linton, and Y.-J. Whang (2009). Nonparametric estimation of a polarization measure. Cemmap Working Papers, CWP14/09.

Anderson, G. J. (1996). Nonparametric tests of stochastic dominance in income distributions. *Econometrica*, 1183–1193.

Angrist, J. and G. M. Kuersteiner (2004). Semiparametric causality tests using the policy propensity score. NBER Working Papers, No. 10975.

Angrist, J. and G. M. Kuersteiner (2008). Causal effects of monetary shocks: Semiparametric conditional independence tests with a multinomial propensity score. IZA Discussion Papers, No. 3606.

Barrett, G. and S. Donald (2003). Consistent tests for stochastic dominance. *Econometrica 71*, 71–104.

Beirlant, J. and D. M. Mason (1995). On the asymptotic normality of L_p-norms of empirical functionals. *Mathematical Methods of Statistics 4*, 1–19.

Berger, E. (1991). Marjorization, exponential inequalities and almost sure behavior of vector-valued random variables. *Annals of Probability 19*, 1206–1226.

Bhattacharya, R. N. (1975). On errors of normal approximation. *Annals of Probability 3*, 815–828.

Bitler, M. P., J. B. G. Gelbach, and H. W. Hoynes (2006). What mean impacts miss: distributional effects of welfare reform experiments. *American Economic Review 96*(4), 988–1012.

Bitler, M. P., J. B. G. Gelbach, and H. W. Hoynes (2007). Distributional impacts of the self-sufficiency project. *Journal of Public Economics 92*(3-4), 748–765.

Blundell, R. and M. Costa Dias (2008). Alternative approaches to evaluation in empirical microeconomics. Cemmap Working Papers, CWP26/08.

Card, D. and D. R. Hyslop (2005). Estimating the effects of a time-limited earnings subsidy for welfare-leavers. *Econometrica 73*(6), 1723–1770.

Chattopadhyay, R. and E. Duflo (2004). Women as policy makers: Evidence from a randomized policy experiment in india. *Econometrica 72*(5), 1409–1443.

Chaudhuri, P. (1991). Nonparametric estimates of regression quantiles and their local bahadur representation. *Annals of Statistics 19*(2), 760–777.

Chaudhuri, P., K. Doksum, and A. Samarov (1997). On average derivative quantile regression. *Annals of Statistics 25*(2), 715–744.

Chen, X., H. Hong, and A. Tarozzi (2008). Semiparametric efficiency in gmm models with auxiliary data. *Annals of Statistics 36*(2), 808–843.

Chernozhukov, V., I. Fernández-Val, and B. Melly (2009). Inference on counterfactual distributions. Cemmap Working Papers, CWP09/09.

Chernozhukov, V. and C. Hansen (2005). An IV model of quantile treatment effects. *Econometrica*, 245–261.

Chernozhukov, V., S. Lee, and A. M. Rosen (2009). Intersection bounds: Estimation and inference. Cemmap Working Papers, CWP19/09.

Crump, R. K., V. J. Hotz, G. W. Imbens, and O. A. Mitnik (2008). Nonparametric tests for treatment effect heterogeneity. *Review of Economics and Statistics 90*(3), 389–405.

Cuevas, A. and R. Fraiman (1997). A plug-in approach to support estimation. *Annals of Statistics 25*, 2300–2312.

Davidson, R. and J.-Y. Duclos (1997). Statistical inference for measurement of the incidence of taxes and transfers. *Econometrica 65*, 1453–1465.

Davidson, R. and J.-Y. Duclos (2000). Statistical inference for stochastic dominance and measurement for the poverty and inequality. *Econometrica 68*, 1435–1464.

Dehejia, R. H. and S. Wahba (1999). Causal effects in nonexperimental studies: Reevaluating the evaluation of training programs. *Journal of the American Statistical Association 94*(448), 1053–1062.

Dehejia, R. H. and S. Wahba (2002). Propensity score-matching methods for nonexperimental causal studies. *Review of Economics and Statistics 84*(1), 151–161.

Delgado, M. A. and W. González Manteiga (2001). Significance testing in nonparametric regression based on the bootstrap. *Annals of Statistics 29*, 1469–1507.

Duflo, E., R. Glennerster, and M. Kremer (2007). Using randomization in development economics research: A toolkit. In T. P. Schults and J. Strauss (Eds.), *Handbook of Development Economics*, Volume 4, pp. 3895–62. Elsevier.

Fan, J. and I. Gijbels (1996). *Local Polynomial Modelling and Its Applications.* London, UK: Chapman & Hall.

Fan, J., T. Hu, and Y. Truong (1994). Robust non-parametric function estimation. *Scandinavian Journal of Statistics 21*, 433–446.

Firpo, S. (2007). Efficient semiparametric estimation of quantile treatment effects. *Econometrica*, 259–276.

Gao, J. and I. Gijbels (2008). Bandwidth selection in nonparametric kernel testing. *Journal of the American Statistical Association 103*(484), 1584–1594.

Ghosal, S., A. Sen, and A. W. van der Vaart (2000). Testing monotonicity of regression. *Annals of Statistics 28*, 1054–1082.

Giné, E., D. M. Mason, and A. Y. Zaitsev (2003). The L_1-norm density estimator process. *Annals of Probability 31*, 719–768.

Glenn, H. and J. A. List (2004). Field experiments. *Journal of Economic Literature*, 1013–1059.

Guerre, E., I. Perrigne, and Q. Vuong (2009). Nonparametric identification of risk aversion in first-price auctions under exclusion resrictions. *Econometrica 77*(4), 1193–1227.

Hahn, J. (1998). On the role of the propensity score in efficient semiparametric estimation of average treatment effects. *Econometrica 66*(2), 315–331.

Heckman, James, H. I. and P. Todd (1998). Matching as an econometric evaluation estimator. *Review of Economic Studies 65*, 261–294.

Heckman, J. J. and E. J. Vytlacil (1999). Local instrumental variable and latent variable models for identifying and bounding treatment effects. *Proceedings of the National Academy of Sciences 96*, 4730–4734.

Heckman, J. J. and E. J. Vytlacil (2005). Structural equations, treatment, effects and econometric policy evaluation. *Econometrica 73*(3), 669–738.

Heckman, J. J. and E. J. Vytlacil (2007a). Econometric evaluation of social programs, part I: Causal models, structural models and econometric policy evaluation. In J. J. Heckman and E. E. Leamer (Eds.), *Handbook of Econometrics*, Volume 6, Chapter 70, pp. 4779–4874. Elsevier.

Heckman, J. J. and E. J. Vytlacil (2007b). Econometric evaluation of social programs, part II: Using the marginal treatment effect to organize alternative econometric estimators to evaluate social programs, and to forecast their effects in new environments. In J. J. Heckman and E. E. Leamer (Eds.), *Handbook of Econometrics*, Volume 6, Chapter 71, pp. 4875–5143. Elsevier.

Hirano, K., G. W. Imbens, and G. Ridder (2003). Efficient estimation of average treatment effects using the estimated propensity score. *Econometrica 71*, 1161–1189.

Hoderlein, S. and E. Mammen (2009). Identification and estimation of local average derivatives in non-separable models without monotonicity. *Econometrics Journal 12*(1), 1–25.

Horváth, L., P. Kokoszka, and R. Zikitis (2006). Testing for stochastic dominance using the weighted McFadden-type statistic. *Journal of Econometrics 133*, 191–205.

Imbens, G. W. (2004). Nonparametric estimation of average treatment effects under exogeneity: A review. *Review of Economics and Statistics 86*, 4–29.

Imbens, G. W. and J. D. Angrist (1994). Identification and estimation of local average treatment effects. *Econometrica 62*(2), 467–475.

Imbens, G. W. and J. Wooldridge (2009). Recent developments in the econometrics of program evaluation. *Journal of Economic Literature 47*(1), 5–86.

Kaur, A., B. L. S. Prakasa Rao, and H. Singh (1994). Testing for second order stochastic dominance of two distributions. *Econometric Theory 10*, 849–866.

Klecan, L., R. McFadden, and D. McFadden (1991). A robust test for stochastic dominance. unpublished working paper.

Kling, J. R., J. B. Liebman, and L. F. Katz (2007). Experimental analysis of neighborhood effects. *Econometrica 75*(1), 83–119.

LaLonde, R. J. (1986). Evaluating the econometric evaluations of training programs with experimental data. *American Economic Review 76*(4), 604–620.

Lavergne, P. (2001). An equality test across nonparametric regressions. *Journal of Econometrics 103*(1-2), 307–344.

Lee, M.-J. (2009). Non-parametric tests for distributional treatment effect for randomly censored responses. *Journal of the Royal Statistical Society: Series B 71*(1), 243–264.

Linton, O. and P. Gozalo (1997). Conditional independence restrictions: testing and estimation. Cowles Foundation Discussion Paper, No. 1140.

Linton, O., E. Maasoumi, and Y.-J. Whang (2005). Consistent testing for stochastic dominance under general sampling schemes. *Review of Economic Studies 72*, 735–765.

Linton, O., K. Song, and Y.-J. Whang (2010). An improved bootstrap test of stochastic dominance. *Journal of Econometrics 154*(2), 186–202.

Manski, C. (2004). Statistical treatment rules for heterogeneous populations. *Econometrica 72*(4), 1221–1246.

Mason, D. M. and W. Polonik (2009). Asymptotic normality of plug-in level set estimates. *Annals of Applied Probability 19*(3), 1108–1142.

McFadden, D. (1989). Testing for stochastic dominance. In T. B. Fomby and T. K. Seo (Eds.), *Studies in the Economics of Uncertainty: In Honor of Josef Hadar*. Springer.

Miguel, E. and M. Kremer (2004). Worms: Identifying impacts on education and health in the presence of treatment externalities. *Statistics and Probability Letters 72*(1), 159–217.

Pinelis, I. F. (1994). On a majorization inequality for sums of independent random variables. *Statistics and Probability Letters 19*, 97–99.

Pollard, D. (1984). *Convergence of Stochastic Processes*. New York, NY: Springer-Verlag.

Shergin, V. V. (1990). The central limit theorem for finitely dependent random variables. In B. Grigelionis, Y. V. Prohorov, V. V. Sazonov, and S. V. (Eds.), *Proc. 5th Vilnius Conf. Probability and Mathematical Statistics*, Volume II, pp. 424–431.

Smith, J. A. and P. E. Todd (2005). Does matching overcome LaLonde's critique of nonexperimental estimators? *Journal of Econometrics 125*(1-2), 305 – 353.

Song, K. (2007). Testing conditional independence via Rosenblatt transforms. *Annals of Statistics*. forthcoming.

Su, L. and H. White (2004). Testing conditional independence via empirical likelihood. Discussion Paper, University of California San Diego.

Su, L. and H. White (2007). A characteristic-function-based test for conditional independence. *Journal of Econometrics 141*, 807–834.

Su, L. and H. White (2008). A nonparametric Hellinger metric test for conditional independence. *Econometric Theory 24*, 829–864.

van der Vaart, A. W. and J. A. Wellner (1996). *Weak Convergence and Empirical Processes*. New York, NY: Springer-Verlag.

Department of Economics, University College London, Gower Street, London, WC1E 6BT, UK.
E-mail address: l.simon@ucl.ac.uk
URL: http://www.homepages.ucl.ac.uk/~uctplso.

Department of Economics, Seoul National University, Seoul 151-742, Korea.
E-mail address: whang@snu.ac.kr
URL: http://plaza.snu.ac.kr/~whang.